

**ALINE RODRIGUES**

**“IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES E  
FATORES RELACIONADOS À FLOCULAÇÃO E  
DESENVOLVIMENTO DE UMA LEVEDURA INDUSTRIAL  
NÃO FLOCULANTE”**

***“IDENTIFICATION AND CHARACTERIZATION OF GENES  
AND FACTORS RELATED TO FLOCCULATION AND  
DEVELOPMENT OF A NON-FLOCCULENT INDUSTRIAL  
YEAST STRAIN”***

**CAMPINAS  
2013**



**UNIVERSIDADE ESTADUAL DE CAMPINAS**  
**INSTITUTO DE BIOLOGIA**

**ALINE RODRIGUES**

**“IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES E  
FATORES RELACIONADOS À FLOCULAÇÃO E  
DESENVOLVIMENTO DE UMA LEVEDURA INDUSTRIAL  
NÃO FLOCULANTE”**

Orientador: Prof. Dr. Gonçalo Amarante Guimarães Pereira

Co-orientador: Prof. Dr. Juan Lucas Argueso



***“IDENTIFICATION AND CHARACTERIZATION OF GENES  
AND FACTORS RELATED TO FLOCCULATION AND  
DEVELOPMENT OF A NON-FLOCCULENT INDUSTRIAL  
YEAST STRAIN”***

Tese apresentada ao Instituto de Biologia para  
obtenção do Título de Doutora em Genética e Biologia Molecular, na área de Genética de  
Microorganismos.

*Thesis presented to the Institute of Biology to  
obtain the Ph.D. grade in Genetics and Molecular Biology, in the field of Genetics of  
Microorganisms.*

**CAMPINAS,  
2013**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Biologia  
Mara Janaina de Oliveira - CRB 8/6972

R618i Rodrigues, Aline, 1983-  
Identificação e caracterização de genes e fatores relacionados à floculação e desenvolvimento de uma levedura industrial não floculante / Aline Rodrigues. – Campinas, SP : [s.n.], 2013.

Orientador: Gonçalo Amarante Guimarães Pereira.

Coorientador: Juan Lucas Argueso.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Leveduras. 2. *Saccharomyces cerevisiae*. 3. Bioetanol. 4. Floculação. 5. Morfologia de colônia. I. Pereira, Gonçalo Amarante Guimarães, 1964-. II. Argueso, Juan Lucas, 1972-. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Identification and characterization of genes and factors related to flocculation and development of a non-flocculent industrial yeast strain

**Palavras-chave em inglês:**

Yeasts

*Saccharomyces cerevisiae*

Bioethanol

Flocculation

Colony morphology

**Área de concentração:** Genética de Microorganismos

**Titulação:** Doutora em Genética e Biologia Molecular

**Banca examinadora:**

Gonçalo Amarante Guimarães Pereira [Orientador]

Iran Malavazi

Flávio Henrique da Silva

Jörg Kobarg

Marcelo Brocchi

**Data de defesa:** 30-10-2013

**Programa de Pós-Graduação:** Genética e Biologia Molecular



Campinas, 30 de outubro de 2013

**BANCA EXAMINADORA**

Prof. Dr. Gonçalo Amarante Guimarães Pereira (orientador)

  
Assinatura

Prof. Dr. Marcelo Brocchi

  
Assinatura

Prof. Dr. Iran Malavazi

  
Assinatura

Prof. Dr. Flavio Henrique Da Silva

  
Assinatura

Dr. Jörg Kobarg

  
Assinatura


Prof. Dr. Fabio Papes

  
Assinatura

Dr. Roberto Ruller

  
Assinatura

Dra. Valéria Maia Merzel

  
Assinatura



## RESUMO

A floculação é um processo de agregação celular observado em muitas linhagens de *Saccharomyces cerevisiae* que resulta em sedimentação, prejudicando o rendimento fermentativo e a operação de destilarias produtoras de bioetanol. A maioria das linhagens utilizadas por essa indústria, incluindo a linhagem PE-2/JAY270, foi selecionada pela incapacidade de flocular.

O controle genético da floculação é bem caracterizado, vários genes foram associados ao fenômeno, incluindo principalmente a família de genes *FLO*. Previamente observamos que a linhagem JAY291 (haploide derivado de JAY270) não flocula, apesar de possuir o alelo funcional do regulador positivo de floculação *FLO8*, sugerindo que seu fenótipo envolve outro passo dessa via. A análise da progênie F2 do cruzamento de JAY291 com a linhagem de laboratório S288 (*flo8-1*) revelou uma proporção fenotípica de ~1:8 (floculantes/não floculantes), esperada para uma característica controlada por três genes não ligados. Sequenciamos os genomas de oito indivíduos F2 floculantes segregando ~40.000 SNPs existentes entre S288c e JAY291. As sequências foram comparadas entre si para localizarmos regiões que co-segregavam com a floculação. Esse mapeamento preliminar foi refinado por PCR-RFLP, revelando um segundo gene, *FLO1*, que está ausente em JAY291. Para identificar o terceiro gene, analisamos fenotipicamente novos haploides F2 segregantes de um cruzamento entre S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*), dentre os quais os indivíduos *FLO1::HphMX4* foram divididos em dois “bulks” (floculante e não floculante) e sequenciados. As frequências alélicas S288c ou JAY291 nos “bulks” foram interrogadas e validadas estatisticamente, revelando um terceiro gene fortemente associado ao fenótipo: *MSS11* (fator de transcrição envolvido em floculação). Encontramos ainda outros doze loci contendo genes de efeito menor, mas que atuam como modificadores do fenótipo.

Em experimentos independentes, observamos alteração na morfologia de colônia de JAY270 e troca de seu típico fenótipo liso para rugoso, também associado com um grau baixo de floculação. Tais colônias rugosas também são frequentemente encontradas na indústria. Curiosamente, o mesmo fenótipo foi observado cruzando-se haploides específicos derivados de JAY270. Os cruzamentos sugeriram que o fenótipo era controlado um único gene, porém específico de diploides. O mapeamento em genomas de haploides de JAY270 revelou o gene responsável pela alteração de fenótipo: *ACE2*, envolvido com separação celular. JAY270 é heterozigota para uma mutação *frameshift* em *ACE2* (*ACE2/ace2-7A*). O fenótipo rugoso ocorre por recombinação mitótica associada com

perda de heterozigosidade (LOH), produzindo células com duas cópias mutadas desse gene.

Para evitar futuros problemas industriais causados pela sedimentação de linhagens rugosas resultantes de LOH na região de *ACE2*, o defeito detectado *ace2-7A* foi reparado no genoma da JAY270. O diploide FGY050 (*ura3Δ/ura3Δ*), derivado de JAY270 e auxotrófico para uracila, foi utilizado como plataforma. A linhagem desenvolvida passou a ter duas cópias do alelo selvagem de *ACE2* (*ACE2/ACE2*) e portanto é incapaz de se tornar rugosa por recombinação.

Finalmente, em um estudo não relacionado com floculação ou morfologia de colônia, testamos se LOH pode estar relacionada com alguma consequência fenotípica benéfica à JAY270. Dentre onze linhagens LOH testadas em ensaios fermentativos de competição direta com o parental, encontramos uma com ganho de competitividade durante a fermentação.

## ABSTRACT

Flocculation is a process of cellular aggregation observed in several *Saccharomyces cerevisiae* strains which results in sedimentation, impairing the fermentation yield and the operation of Brazilian bioethanol distilleries. Most strains used in this industry, including the PE-2/JAY270 strain, were selected for their inability to flocculate.

The genetic control of flocculation is well characterized and several genes were associated with this phenomenon, including the *FLO* family of genes. We previously observed that the JAY291 strain (derivative haploid of JAY270) does not flocculate, despite having the functional allele of the positive regulator of flocculation *FLO8*, suggesting that its phenotype involves another step of this pathway. Analysis of the F2 progeny from crossing JAY291 with the laboratory strain S288c (*flo8-1*) revealed the phenotypic ratio of ~1:8 (flocculent/non-flocculent), expected for a trait controlled by three unlinked genes. We sequenced the genomes of eight flocculent F2 individuals segregating ~40.000 SNPs that exist between S288c and JAY291. The sequences were compared to each other to localize regions that co-segregated with the flocculation. This preliminary mapping was refined by PCR-RFLP, revealing a second gene, *FLO1*, which is absent in JAY291. To identify the third gene, we phenotypically analyzed new F2 haploids segregating from a cross between S288c (*FLO8::KanMX4 FLO1::HphMX4*) and JAY291 (*FLO8 flo1Δ*), among which *FLO1:HphMX4* individuals were separated into two bulks (flocculent and non-flocculent) and sequenced. Allele frequencies S288c or JAY291 in bulks were interrogated and statistically validated, revealing a third gene strongly associated with the phenotype: *MSS11* (transcription factor involved in flocculation). We also found twelve other loci containing minor effect genes, but acting as modifiers of the phenotype.

In independent experiments, we observed alterations in the colony morphology of JAY270 and a transition from its typical rough to smooth phenotype, also associated with a low degree of flocculation. This type of colony is often found in the industry. Interestingly the same phenotype was observed crossing specific haploids derivatives of JAY270. The crosses suggested that the phenotype was controlled by a single gene, but diploid specific. Mapping genomes of haploids of JAY270 revealed the gene responsible for the altered phenotype: *ACE2*, involved in cell separation. The JAY270 is heterozygous to a frameshift mutation in the *ACE2* (*ACE2/ace2-7A*). The rough phenotype occurs by mitotic recombination associated with loss-of-heterozygosity (LOH), producing cells with two mutated copies of this gene.

To avoid future industrial problems caused by sedimentation of rough strains derived from LOH in the *ACE2* region, the *ace2-7A* defect was repaired in the JAY270 genome. The diploid FGY050 (*ura3Δ/ura3Δ*), derived from JAY270 and auxotrophic to uracil was used as a platform. The developed strain has two copies of the *ACE2* wild-type allele (*ACE2/ACE2*) and is therefore unable to become roughened by recombination.

Finally, in a study unrelated to flocculation or colony morphology, we tested whether LOH may be associated with some phenotypic consequence beneficial to the JAY270. Out of eleven LOH strains tested in fermentative assays of direct competition against their parental strain, we found one with gain in competitiveness during fermentation.

## Sumário

Resumo .....	vii
Abstract .....	ix
Sumário .....	xi
Agradecimentos.....	xix
Estrutura da tese .....	1
Introdução Geral e Objetivos .....	3
Introdução Geral.....	5
Linhagens industriais de <i>Saccharomyces cerevisiae</i> utilizadas para produção de etanol no Brasil .....	5
Floculação em <i>Saccharomyces cerevisiae</i> .....	6
Genes envolvidos com floculação descritos na literatura.....	8
Prejuízos causados por floculação ao processo de produção de etanol.....	11
Objetivos.....	13
Referências .....	14
Capítulo 1.....	17
Mapeamento genômico por novos genes de floculação na linhagem industrial PE-2 (JAY291) .....	17
Introdução .....	19
Materiais e Métodos .....	23
Linhagens, Meios, técnicas moleculares/microbiológicas .....	23
Mapeamento genômico preliminar .....	28
Refinando o mapeamento por PCR-RFLP ( <i>Polymerase Chain Reaction -Restriction Fragment Length Polymorphism</i> ) .....	30
Construção do plasmídeo pAR2 .....	32
Reparo da mutação <i>nonsense flo8-1</i> e inserção do marcador Hph ligado ao gene <i>FLO1</i> no genoma de S288c .....	36
Inserção do marcador Hph ligado ao gene <i>FLO1</i> no genoma de S288c.....	39
Mapeamento genômico em “bulks” - <i>Bulk Segregant Analysis</i> (BSA) aliada a sequenciamento de última geração - <i>Next Generation Sequencing</i> (NGS).....	41
Resultados .....	45
Mapeamento genômico preliminar .....	45
Genotipagem adicional por PCR-RFLP: refinando o mapeamento .....	47
<i>FLO1</i> nos genomas de S888c e JAY291.....	51

Mapeamento genômico em “bulks” .....	53
Construção do plasmídeo pAR2 .....	54
Construção de uma linhagem S288c <i>FLO8</i> .....	54
Inserção de um marcador de seleção ligado ao gene <i>FLO1</i> de S288c .....	56
Nova população F2 segregante para análise.....	59
Mapeamento genômico: Frequência alélica parental X Posição cromossômica (SNP)62	
Loci associados com o fenótipo floculante encontrados pelo mapeamento e validados estatisticamente .....	79
Discussão .....	83
Conclusões.....	87
Referências .....	87
Capítulo 2.....	91
Colony morphology transition in industrial yeast strains is associated with mitotic recombination leading to loss-of-heterozygosity .....	91
Introduction.....	93
Materials and Methods .....	94
Media, growth conditions, strains and microbiological techniques .....	94
Analysis of the <i>ACE2</i> region.....	96
Results.....	98
Observation of rough colony morphology among derivatives of the JAY270 strain ...	98
Identification of the mutation responsible for the rough colony phenotype.....	102
Analysis of the genomic region containing <i>ACE2</i> in spontaneous JAY270-derived rough colony isolates.....	104
Discussion .....	107
PE-2 colony morphology transition is due to a mutation in <i>ACE2</i> , a gene involved in the cell separation process.....	107
The rough morphology only appears in diploid cells and mutations in other genes of the pathway can also lead to the same phenotype as <i>ace2</i> .....	108
The main mechanism leading to PE-2 colony variation is Loss-of-Heterozygosity (LOH) .....	109
Conclusions.....	109
References .....	110
Capítulo 3.....	113
Manipulação genética de linhagem PE-2 para reparo de defeito identificado em gene envolvido com variação de morfologia de colônia .....	113



Introdução .....	115
Materiais e Métodos .....	116
Meios de cultura, condições de crescimento, linhagens e técnicas microbiológicas/moleculares .....	116
Construção de plasmídeo .....	118
Construção de linhagem PE-2 homozigota para o <i>locus</i> selvagem de <i>ACE2</i> .....	121
Resultados .....	124
Estratégia para reposição alélica utilizada para construção de linhagem PE-2 homozigota para <i>locus</i> tipo selvagem de <i>ACE2</i> .....	124
Construção do plasmídeo pAR1 .....	125
Reparo do gene <i>ACE2</i> em uma linhagem PE-2 .....	125
Discussão .....	128
Conclusões .....	129
Referências .....	130
Capítulo 4.....	131
Análise das consequências fenotípicas da perda de heterozigosidade (LOH) durante a fermentação de etanol .....	131
Introdução .....	133
Materiais e Métodos .....	134
Meios de cultura e linhagens de levedura .....	134
Ensaio fermentativos de competição entre linhagens LOH e parental .....	135
Resultados .....	138
Discussão .....	143
Conclusões .....	144
Referências .....	145
Conclusões Gerais.....	147
Anexo I .....	151
A simple and effective set of PCR-based molecular markers for the monitoring of the <i>Saccharomyces cerevisiae</i> cell population during bioethanol fermentation .....	151
Anexo II .....	163



*Dedico este trabalho aos meus pais Sônia (in memoriam) e Paulo,  
aos meus irmãos Érika e Paulo e ao meu noivo André,  
pelo amor e apoio.*



*“The more original a discovery, the  
more obvious it seems afterwards”*

*Arthur Koestler*



## **Agradecimentos**

Agradeço primeiramente aos meus pais, Sônia e Paulo, por todo o apoio que sempre me deram, pelo amor, carinho, amizade, confiança, palavras de incentivo, por sempre terem priorizado a educação de seus filhos e por todos os ensinamentos que estarão guardados para sempre em meu coração.

Aos meus irmãos, Érika e Paulo, pelo companheirismo, amizade, amor e apoio.

Ao meu noivo André, por toda a compreensão e apoio durante um ano de distância para realização de meu doutorado sanduíche, pelo amor, companheirismo e amizade.

Ao meu orientador Prof. Dr. Gonçalo Pereira, pelas oportunidades oferecidas e confiança depositada em meu trabalho.

Ao meu co-orientador Prof. Dr. Lucas Argueso, por toda sua dedicação durante o desenvolvimento de meus estudos, pela ótima recepção que tive durante um ano de doutorado sanduíche realizado em seu laboratório na Colorado State University, pelos ensinamentos, por toda a ajuda, disponibilidade e valiosas discussões durante todos esses anos.

Aos colegas do laboratório do LGE e da CSU, pela ajuda no dia a dia, conversas e momentos de descontração. Em especial, agradeço aos colegas Sílvia, por toda a ajuda oferecida no início de meu doutorado, Ane, pela recepção na minha chegada aos EUA, Theo, Nádia, Chris, Guadalupe, Pedro e Osmar pelos experimentos que realizaram e que serão muito importantes para os artigos resultantes.

Ao Prof. Dr. Paul Magwene, da Duke University, pelas análises estatísticas realizadas nesse trabalho.

À equipe administrativa do LGE, Sílvia, Eliane, Mariana, Ernestina e Welbe, por todo o apoio concedido.

Às agências CAPES e CNPq, pelo auxílio financeiro concedido através de bolsas de pesquisa.





## **Estrutura da tese**

Primeiramente, são apresentados uma introdução geral e objetivos do trabalho. Em seguida, essa tese foi quatro capítulos. Em cada capítulo há uma introdução individual, materiais e métodos, resultados, discussão, conclusões e referências. O capítulo 2 contém dados e experimentos de um estudo em fase final de redação de manuscrito, o qual será submetido à revista *Genetics* em breve. Sou a primeira e principal autora nesse manuscrito. No anexo I é apresentado um manuscrito aceito para publicação na revista *Journal of Biotechnology*, no qual tive participação como co-autora pela realização de alguns experimentos de genotipagem de linhagens por PCR.

### **Introdução Geral e Objetivos**

**Capítulo 1:** Mapeamento genômico por novos genes de flocculação na linhagem industrial PE-2 (JAY291)

**Capítulo 2:** Colony morphology transition in industrial yeast strains is associated with mitotic recombination leading to loss-of-heterozygosity

**Capítulo 3:** Manipulação genética de linhagem PE-2 para reparo de defeito identificado em gene envolvido com variação de morfologia de colônia

**Capítulo 4:** Análise das consequências fenotípicas da perda de heterozigosidade (LOH) durante a fermentação de etanol

**Conclusões Gerais**

**Anexo I:** A simple and effective set of PCR-based molecular markers for the monitoring of the *Saccharomyces cerevisiae* cell population during bioethanol fermentation

## **Introdução Geral e Objetivos**



## INTRODUÇÃO GERAL

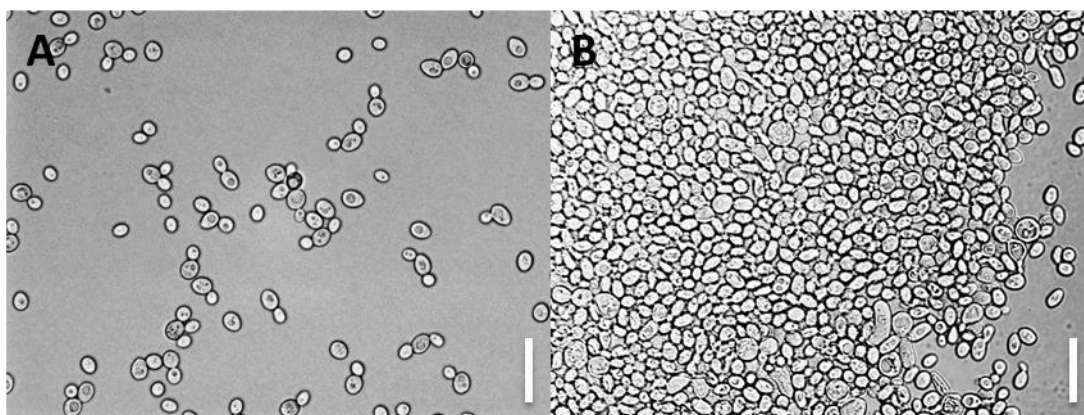
### **Linhagens industriais de *Saccharomyces cerevisiae* utilizadas para produção de etanol no Brasil**

As linhagens industriais *Saccharomyces cerevisiae* utilizadas atualmente para produção de etanol foram isoladas durante o processo fermentativo em destilarias brasileiras e selecionadas após serem caracterizadas como boas fermentadoras e capazes de dominar o processo fermentativo durante toda a safra. Diversas linhagens vêm sendo utilizadas como pré-inóculo em diferentes unidades industriais: BG-1, CR-1, SA-1, CAT-1, PE-2, entre outras (Basso *et al.* 2008).

A linhagem industrial Pedra-2 (PE-2), isolada na Usina da Pedra e amplamente utilizada por destilarias brasileiras, foi selecionada por apresentar muitas características favoráveis ao processo de produção de etanol, como: alta produtividade e rendimento fermentativo, alta taxa de crescimento e viabilidade celular, alta tolerância à temperatura e estresse oxidativo, fermentação estável durante toda a temporada, baixa formação de espuma, extremamente bem sucedida em competições e não floculante (Basso *et al.* 2008).

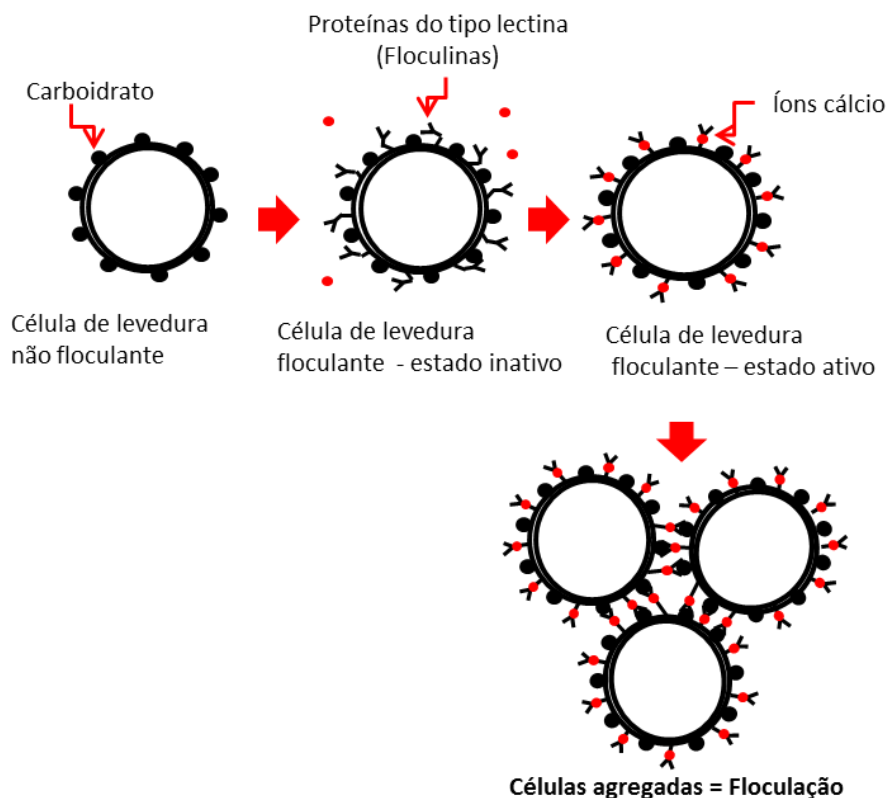
## Floculação em *Saccharomyces cerevisiae*

Floculação é a agregação de células de leveduras em massas multicelulares, compostas por centenas e até milhares de células, chamadas de flocos, que resulta na rápida sedimentação celular (Soares 2011).



**Figura 1 - Fenótipos de linhagem não floculante versus linhagem floculante, observados em microscopia.** Exemplos de células livres e não floculantes (A), em comparação com células floculadas, que formam grandes aglomerados multicelulares (B). Em “A” são células da linhagem industrial PE-2/JAY270 e em “B”, células da linhagem de laboratório S288c/ARY156, com o gene *FLO8* restaurado (desenvolvida no capítulo 2 desse trabalho). A barra indica 20 µm.

O mecanismo físico pelo qual ocorre a floculação é denominado *Lectin-like Theory of Flocculation*, teoria que sugeriu a ligação que ocorre entre lectinas (floculinas) presentes na parede celular de células floculantes e resíduos de açúcar presentes na parede celular de células vizinhas, sendo esse processo reversível e dependente de cálcio (Miki *et al.* 1982; Touhami *et al.* 2003; Singh *et al.* 2011).



**Figura 2 - Teoria de flocculação chamada de *Lectin-like*** (Miki *et al.* 1982). As proteínas de parede celular, chamadas de floculinas, reconhecem carboidratos presentes na parede de células vizinhas. Para atingir sua conformação ativa, as floculinas requerem a presença de íons cálcio. A agregação de células resultante dessa interação é chamada de flocculação (adaptado de Singh *et al.* 2011).

As floculinas possuem uma estrutura comum de três domínios: um domínio C-terminal, que contém uma âncora-GPI (glycosylphosphatidylinositol) que liga a floculina à parede celular; um domínio N-terminal, que sobressai da superfície celular e se liga a um carboidrato presente na parede celular de uma célula vizinha e um amplo domínio mediano caracterizado por múltiplas repetições ricas em serina e treonina codificadas por sequências de DNA conservadas (Verstrepen and Klis 2006).

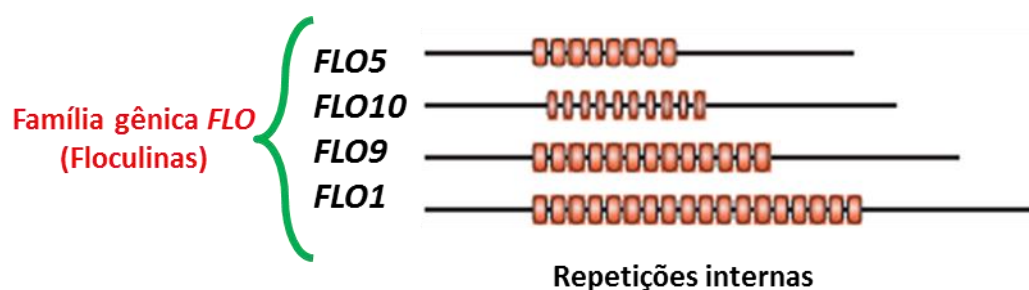
A floculação pode ocorrer em condições de estresse, comuns aos processos fermentativos industriais, como diminuição da quantidade de nutrientes, alta concentração de etanol, condições de temperatura, pH e oxigênio não adequados ao crescimento e interação com bactérias presentes nas dornas de fermentação (*Lactobacillus fermentum*) (Teresinha and Fumio 1993; Bromberg and Yokoya 1995; Soares and Seynaeve 2000; Sampermans *et al.* 2005; Smukalla *et al.* 2008; Soares 2011).

### **Genes envolvidos com floculação descritos na literatura**

Em *S. cerevisiae*, a floculação é controlada geneticamente e muitos genes foram relatados como sendo responsáveis por este fenômeno, sendo principalmente pertencentes à família *FLO*. Entre eles estão: *FLO1*, *FLO5*, *FLO9* e *FLO10* (Teunissen and Steensma 1995; Soares 2011).

Os genes *FLO* possuem muitas repetições internas *in tandem*. Essas unidades internas de repetições favorecem eventos de recombinação, resultando na expansão ou contração do tamanho do gene. O número dessas unidades de repetição foi também relacionado com o aumento ou diminuição no grau de floculação. O *FLO1* é o que possui o maior número de repetições internas entre os genes da família *FLO* e foi observado que a expansão de repetições em *FLO1* resulta em um grau de floculação mais forte. (Verstrepen *et al.* 2005) (Figura 3). Possivelmente, floculinas menores conferem um menor grau de floculação, pois o domínio N-terminal permanece dentro da parede celular (Verstrepen and Klis 2006).





**Figura 3 – Subunidades de repetições intragênicas *in tandem* encontradas em genes da família *FLO*.** O gene *FLO1* possui o maior número de repetições internas dentre os genes dessa família. O aumento ou diminuição no número de repetições causam a expansão ou contração no tamanho do gene. A adição ou remoção dessas unidades de repetição não alteram o quadro de leitura. As subunidades intragênicas de repetição estão demonstradas na figura em cada uma das caixas ou retângulos vermelhos (Verstrepen *et al.* 2005).

*FLO1* é o principal agente no processo de floculação (Kobayashi *et al.* 1998; Smukalla *et al.* 2008). O domínio de ligação N-terminal de Flo1p se liga especificamente a carboidratos de manose (Goossens *et al.* 2011).

O estudo realizado por Smukalla e colaboradores (2008) demonstrou que a expressão de *FLO1* protege células floculantes de múltiplos estresses, principalmente químicos, incluindo altas concentrações de etanol, do antifúngico Anfotericina B (droga comumente utilizada para combater fungos patogênicos como *Candida albicans*) e de peróxido (estresse oxidativo). Ainda nesse estudo, foi demonstrado que as células floculantes presentes no interior dos flocos estão fisicamente protegidas e são mais resistentes do que células presentes nas camadas exteriores dos flocos.

Porém, foi também observado que a expressão de *FLO1* tem o custo de um crescimento mais lento, relativamente às células que não expressam *FLO1*, provavelmente pelo acesso limitado aos nutrientes e oxigênio em células presentes no interior dos flocos. Mesmo quando a floculação foi inibida com adição de manose, as células que expressam

*FLO1* continuaram a apresentar esse defeito no crescimento. Para evitar que células que não expressam *FLO1* se aproveitem dessa proteção, células *FLO1*<sup>+</sup> preferencialmente se aderem entre si, independentemente do parentesco genético do restante do genoma (por exemplo, entre duas espécies diferentes). Por esse motivo, os autores desse trabalho indicaram *FLO1* como um “gene egoísta” (idéia popularizada por Dawkins, 1976) de “efeito barba verde” (*green beard gene*), direcionando a cooperação para outros portadores do mesmo gene. Também foi demonstrado que essa resistência ao estresse aumenta com o aumento no número das repetições *in tandem* do gene *FLO1* (Smukalla *et al.* 2008).

Existem também outros genes, reguladores transcricionais, atuantes sobre a floculação e descritos na literatura. Tup1p, Ssn8p, Srb8p, Sin4p atuam como repressores transcricionais de *FLO1*. Também atuam como repressores de outro gene da família *FLO*, *FLO11* (Fichtner *et al.* 2007), que está relacionado principalmente com fenótipos resultantes de adesão célula-substrato, observada em biofilmes (Reynolds and Fink 2001).

*MSS11*, que participa na regulação da degradação do amido, na formação de pseudohifas e crescimento invasivo haplóide, foi identificado como um forte indutor da floculação (Bester *et al.* 2006).

Foi observado que a superexpressão de outro gene, *GTS-1*, em uma linhagem *Saccharomyces cerevisiae* de laboratório, W303-1A, que possui uma mutação no gene *FLO8* tornando-o não funcional, leva a um fenótipo floculante. A floculação nesse caso era dependente da expressão de *FLO1* (Shen *et al.* 2006).

O gene *FLO8* codifica um ativador do gene *FLO1* (Kobayashi *et al.* 1998). Regula características de crescimento invasivo, produção de glicoamilase extracelular e floculação através da regulação dos genes *FLO11*, *STA1* e *FLO1* (Kobayashi *et al.* 1999).

A linhagem padrão de laboratório S288c contém uma mutação no gene *FLO8* (*flo8-1*) que torna as células não floculantes (Liu *et al.* 1996). A reposição do gene *FLO8* por uma versão intacta e a superexpressão desse gene permite a expressão dos genes *FLO1* e *FLO11* e consequentemente restaura em a floculação e o crescimento pseudohifas, respectivamente, em diplóides (Bester *et al.* 2006; Fichtner *et al.* 2007). Em JAY291, uma linhagem haplóide não floculante derivada de um diplóide PE-2 (isolado JAY270) que teve seu genoma sequenciado recentemente (Argueso *et al.* 2009), o gene *FLO8* aparece como funcional, indicando que o fenótipo não floculante envolve outros genes dessa via. *FLO8* atua como regulador positivo tanto do gene *FLO1* como de *FLO11* (Fichtner *et al.* 2007).

Como já mencionado anteriormente, os genes *FLO* participam de outros tipos de fenótipos relacionados não só à adesão célula-célula (floculação), como também à adesão célula-substrato (*FLO11*) como crescimento invasivo, formação de pseudohifas e biofilmes (Reynolds and Fink 2001; Bauer *et al.* 2010)

### **Prejuízos causados por floculação ao processo de produção de etanol**

A floculação prejudica o processo de produção de etanol, por reduzir o contato da levedura com o meio, resultando em açúcar residual e reduzindo a produtividade fermentativa. Pode ainda diminuir a eficiência de centrifugação, dificultando o tratamento ácido utilizado para reduzir contaminantes bacterianos no final de cada ciclo

fermentativo, resultando no aumento do uso de ácido sulfúrico, aumentando o estresse causado às leveduras, que podem reduzir suas atividades ou morrer. Além disso, pode causar o entupimento de tubulações nas centrífugas utilizadas, causando sérios prejuízos ao processo (Basso *et al.* 2008; Amorim *et al.* 2011).

Diferentes estudos envolvem o processo de adesão celular: patogenicidade de vírus, fungos e bactérias e formação de biofilmes em microbiologia e, abrangendo as áreas, comunicação celular e metástase tumoral (Goossens and Willaert 2010). Considerando que a floculação é um processo de adesão celular ainda não totalmente compreendido e um dos grandes problemas encontrados nas usinas brasileiras produtoras de etanol combustível, a identificação e a compreensão de fatores genéticos envolvidos neste processo podem ter fundamental importância para diversas áreas e poderão ainda, servir como ferramenta para o desenvolvimento de linhagens industriais geneticamente melhoradas para este fenômeno.

## OBJETIVOS

O presente projeto objetiva compreender os fatores genéticos envolvidos no processo de floculação em *Saccharomyces cerevisiae* e revelar novos reguladores desse processo.

Os objetivos específicos deste projeto são:

- ✓ Compreender os fatores genéticos que controlam o fenótipo não floculante da linhagem industrial JAY291 (isolado da PE-2);
- ✓ Identificar genes envolvidos no processo de floculação através de mapeamento genômico;
- ✓ Caracterizar funcionalmente genes selecionados através de deleção gênica e/ou substituição alélica;
- ✓ Obter uma linhagem industrial não floculante, independente das condições encontradas nas dornas de fermentação, com características de interesse industrial preservadas.

## REFERÊNCIAS

- Amorim, H. V., M. L. Lopes, J. V. de Castro Oliveira, M. S. Buckeridge, and G. H. Goldman, 2011 Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 1267–1275.
- Argueso, J. L., M. F. Carazzolle, P. A. Mieczkowski, F. M. Duarte, O. V. C. Netto *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19: 2258–70.
- Basso, L. C., H. V. de Amorim, A. J. de Oliveira, and M. L. Lopes, 2008 Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res.* 8: 1155–63.
- Bauer, F. F., P. Govender, and M. C. Bester, 2010 Yeast flocculation and its biotechnological relevance. *Appl. Microbiol. Biotechnol.* 88: 31–9.
- Bester, M. C., I. S. Pretorius, and F. F. Bauer, 2006 The regulation of *Saccharomyces cerevisiae* *FLO* gene expression and Ca<sup>2+</sup>-dependent flocculation by Flo8p and Mss11p. *Curr. Genet.* 49: 375–83.
- Bromberg, R., and F. Yokoya, 1995 Chemical modifications of the cell-surface components of *Lactobacillus fermentum* FTPT 1405 and their effect on the flocculation of *Saccharomyces cerevisiae*. *World J. Microb. Biot.* 508–511.
- Fichtner, L., F. Schulze, and G. H. Braus, 2007 Differential Flo8p-dependent regulation of *FLO1* and *FLO11* for cell-cell and cell-substrate adherence of *S. cerevisiae* S288c. *Mol. Microbiol.* 66: 1276–89.
- Goossens, K. V. Y., C. Stassen, I. Stals, D. S. Donohue, B. Devreese *et al.*, 2011 The N-terminal domain of the Flo1 flocculation protein from *Saccharomyces cerevisiae* binds specifically to mannose carbohydrates. *Eukaryot. Cell* 10: 110–7.
- Goossens, K., and R. Willaert, 2010 Flocculation protein structure and cell-cell adhesion mechanism in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 32: 1571–85.
- Kobayashi, O., N. Hayashi, R. Kuroki, and H. Sone, 1998 Region of *FLO1* proteins responsible for sugar recognition. *J. Bacteriol.* 180: 6503–10.
- Kobayashi, O., H. Yoshimoto, and H. Sone, 1999 Analysis of the genes activated by the *FLO8* gene in *Saccharomyces cerevisiae*. *Curr. Genet.* 36: 256–61.
- Liu, H., C. A. Styles, and G. R. Fink, 1996 *Saccharomyces cerevisiae* S288C has a mutation in *FLO8*, a gene required for filamentous growth. *Genetics* 144: 967–978.

- Miki, B. L., N. H. Poon, A. P. James, and V. L. Seligy, 1982 Possible mechanism for flocculation interactions governed by gene *FLO1* in *Saccharomyces cerevisiae*. J. Bacteriol. 150: 878–89.
- Reynolds, T. B., and G. R. Fink, 2001 Bakers' yeast, a model for fungal biofilm formation. Science 291: 878–81.
- Sampermans, S., J. Mortier, and E. V Soares, 2005 Flocculation onset in *Saccharomyces cerevisiae*: the role of nutrients. J. Appl. Microbiol. 98: 525–31.
- Shen, H., H. Iha, S. Yaguchi, and K. Tsurugi, 2006 The mechanism by which overexpression of Gts1p induces flocculation in a *FLO8*-inactive strain of the yeast *Saccharomyces cerevisiae*. FEMS Yeast Res. 6: 914–23.
- Singh, R. S., R. Bhari, and H. P. Kaur, 2011 Characteristics of yeast lectins and their role in cell-cell interactions. Biotechnol. Adv. 29: 726–31.
- Smukalla, S., M. Caldara, N. Pochet, A. Beauvais, C. Yan *et al.*, 2008 *FLO1* is a variable green beard gene that drives biofilm-like cooperation in budding yeast. Cell 135: 726–737.
- Soares, E. V, 2011 Flocculation in *Saccharomyces cerevisiae*: a review. J. Appl. Microbiol. 110: 1–18.
- Soares, E. V, and J. Seynaeve, 2000 Induction of flocculation of brewer's yeast strains of *Saccharomyces cerevisiae* by changing the calcium concentration and pH of culture medium. Biotechnol. Lett. 1827–1832.
- Teresinha, M., and S. A. D. Fumio, 1993 Characteristics of yeast cell flocculation by *Lactobacillus fermentum*. J. Ferment. Bioeng. 75: 151–154.
- Teunissen, A. W. R. H., and H. Y. Steensma, 1995 The dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. Yeast 11: 1001–1013.
- Touhami, A., B. Hoffmann, A. Vasella, F. A. Denis, and Y. F. Dufrêne, 2003 Aggregation of yeast cells: direct measurement of discrete lectin-carbohydrate interactions. Microbiology 149: 2873–2878.
- Verstrepen, K. J., A. Jansen, F. Lewitter, and G. R. Fink, 2005 Intragenic tandem repeats generate functional variability. Nat. Genet. 37: 986–90.
- Verstrepen, K. J., and F. M. Klis, 2006 Flocculation, adhesion and biofilm formation in yeasts. Mol. Microbiol. 60: 5–15.





## **Capítulo 1**

### **Mapeamento genômico por novos genes de floculação na linhagem industrial PE-2 (JAY291)**



## **Mapeamento genômico por novos genes de floculação na linhagem industrial PE-2 (JAY291)**

### **INTRODUÇÃO**

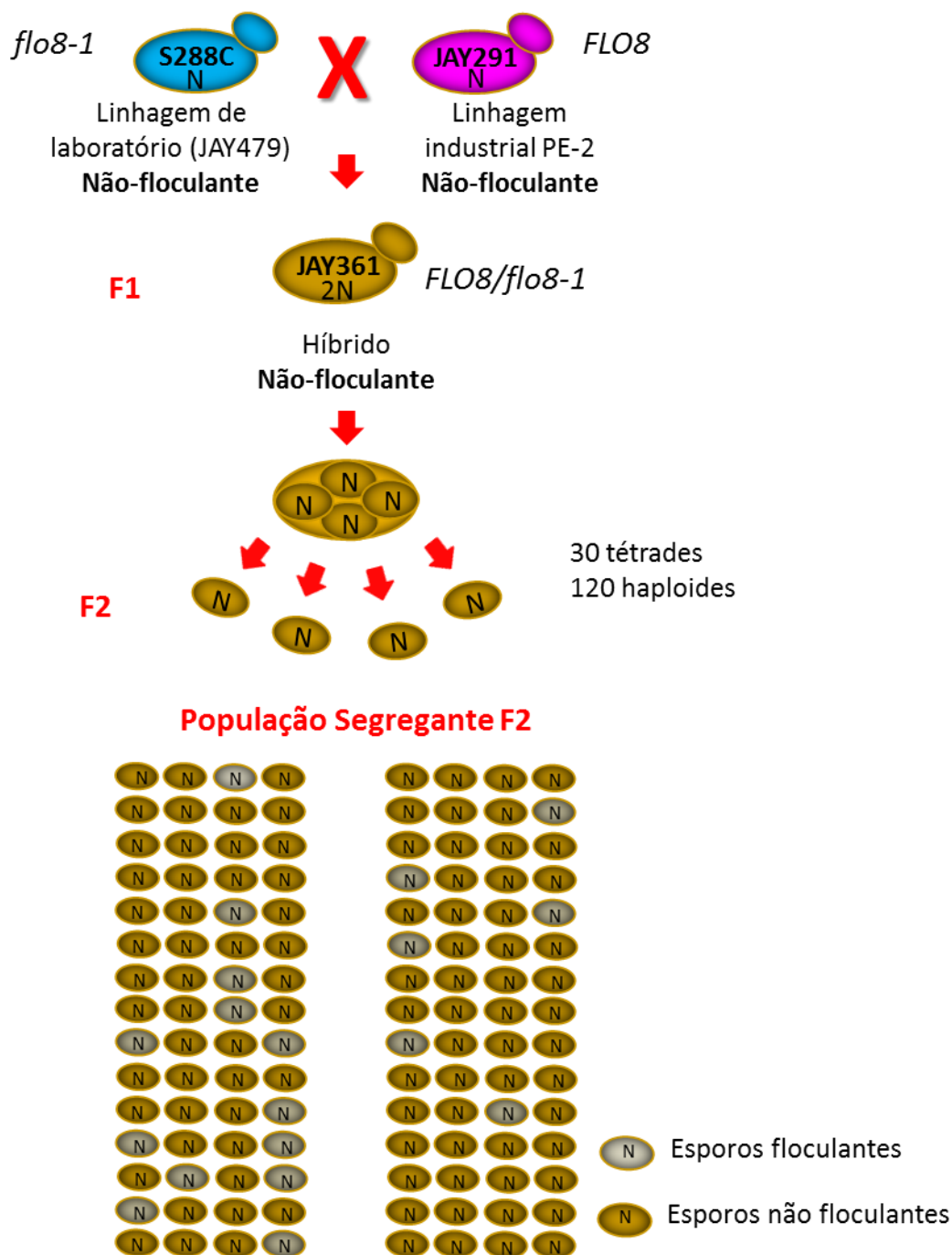
*Saccharomyces cerevisiae* é considerado um modelo para estudos genômicos, principalmente por possuir características como curto tempo de geração sexual, alta taxa de recombinação meiótica e pequeno tamanho de seu genoma (Connelly and Akey 2012; Liti and Louis 2012). Foi o primeiro organismo eucarioto (linhagem de laboratório S288c) a ter seu genoma sequenciado (Goffeau *et al.* 1996). É também um importante organismo produtor na indústria biotecnológica (revisado em Carvalho-Netto, Rodrigues *et al.* 2012).

A maioria das características de interesse industrial de *S. cerevisiae* são quantitativas, ou seja, apresentam uma distribuição contínua de um caráter mensurável e são controladas por múltiplos loci, chamados de QTL (*Quantitative Trait Loci*). Por outro lado, características qualitativas são referidas como Mendelianas, controladas por um único locus de efeito discreto e podem ser consideradas como fenótipos extremos de características quantitativas. Acredita-se ainda que cada um dos numerosos loci que controlam as características quantitativas exerce um pequeno efeito sobre o fenótipo, chamados de loci de efeito menor (*minor effect genes*), mas alguns deles possuem amplo efeito e esses são chamados de loci de efeito maior (*major effect genes*) (Swinnen *et al.* 2012).

Em *S. cerevisiae*, o mapeamento de QTL é normalmente realizado após o cruzamento de duas linhagens parentais que diferem quanto à característica em estudo. Os indivíduos segregantes, após esporulação do híbrido, diferem geneticamente e são

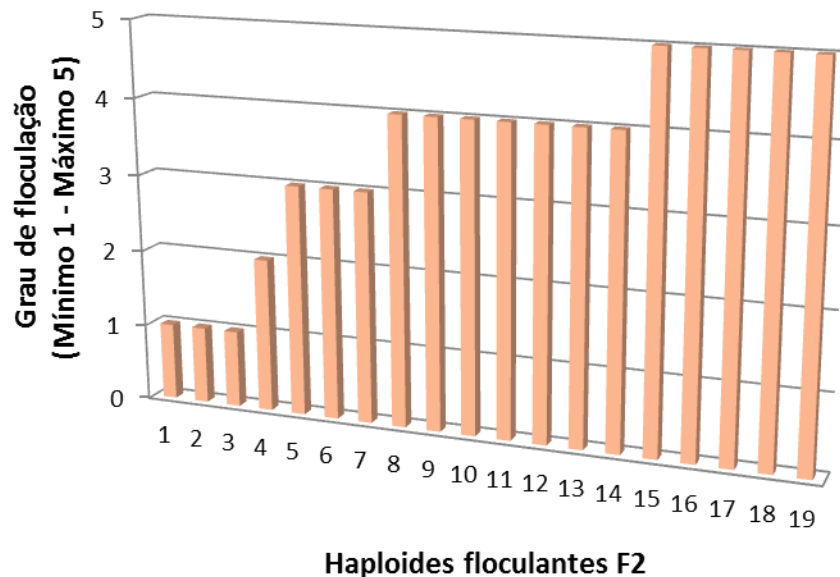
analisados quanto à expressão fenotípica da característica quantitativa de interesse. A variação fenotípica entre as linhagens parentais não é uma necessidade, mas deve existir na população de segregantes (Mancera *et al.* 2008; Swinnen *et al.* 2012).

Em estudos preliminares realizados por nosso grupo (Argueso *et al.* 2009), foi realizado um cruzamento entre a linhagem de laboratório não floculante S288c (JAY479) com a linhagem PE-2 industrial não floculante JAY291, resultando em um híbrido, chamado JAY361, que apresentou o mesmo fenótipo não floculante de seus parentais. Foram observados os fenótipos 120 indivíduos haploides, segregantes da população F2, resultantes da dissecação de trinta tétrades completas (com seus quatro esporos viáveis) do híbrido JAY361. A observação do crescimento em meio líquido de cada um dos esporos separadamente mostrou que dezenove indivíduos apresentavam floculação e 101 eram não floculantes, proporção fenotípica aproximada de 1:8 floculantes para não floculantes, esperada para uma característica controlada por três genes não ligados ( $\chi^2 = 1.22$ ;  $P = 0.27$ ) (Figura 1).



**Figura 1 - Cruzamento da linhagem de laboratório S288c (JAY479) com a linhagem industrial PE-2 (JAY291).** O cruzamento entre S288c (*flo8-1*) e JAY291 (*FLO8*), ambas não floculantes, resulta no híbrido diploide JAY361 (*FLO8/flo8-1*) que apresenta mesmo fenótipo dos parentais. Dentre 120 indivíduos F2 segregantes, 101 não flocularam e 19 flocularam em meio YPD líquido, uma proporção fenotípica de ~1:8 esperada para uma característica controlada por três genes não ligados ( $\chi^2 = 1.22$ ;  $P = 0.27$ ).

Dentre os indivíduos floculantes, observamos que havia uma variável no grau de floculação. Alguns sedimentavam para o fundo do tubo imediatamente após a retirada do vórtex, enquanto que alguns indivíduos também apresentavam flocos visíveis a olho nu no meio líquido, mas demoravam um pouco mais para sedimentar. Por comparação entre o tempo de sedimentação apresentado pelos dezenove indivíduos, os classificamos como nível cinco para a máxima floculação (menor tempo para sedimentação), nível um para floculação mínima (maior tempo para sedimentação dentre os floculantes), sendo nível zero para não floculantes. O gráfico representado na Figura 2 mostra o grau de floculação observado nos dezenove indivíduos floculantes.



**Figura 2 - Grau de floculação entre os dezenove haploides floculantes F2.** Os dezenove indivíduos floculantes foram comparados entre si, quanto ao grau de floculação, pelo tempo de sedimentação em meio YPD líquido. Quando as linhagens sedimentavam imediatamente após retirada do tudo do vórtex, a linhagem foi pontuada como 5 (máximo grau de floculação). Para linhagens não floculantes o grau é zero.

Nesse capítulo, realizamos um mapeamento genômico com intuito de localizarmos regiões que co-segregam com o fenótipo de floculação e controlam o fenótipo não

floculante de JAY291. Observando os primeiros resultados descritos acima, levantamos uma hipótese para o modelo de herança genética de floculação: existe um componente discreto atuante, que determina se a linhagem é flocculante ou não; por outro lado, existe um componente quantitativo, que determina o grau de floculação. Os três genes não ligados, esperados pela proporção fenotípica de 1:8 observada na população F2, poderiam ser os atuantes do componente discreto da herança. Para o componente quantitativo, esperamos encontrar outros genes, que seriam os genes de menor efeito (*minor effect genes*) de uma herança quantitativa, mas também atuantes sobre a característica.

## MATERIAIS E MÉTODOS

### Linhagens, Meios, técnicas moleculares/microbiológicas

As linhagens de levedura utilizadas nesse trabalho são todas espécies de *Saccharomyces cerevisiae* e estão listadas na Tabela 1. A linhagem *Escherichia coli* ( $\alpha$ -*Select Chemically Competent Cells*, Bioline) foi utilizada para clonagem de plasmídeos de acordo com as instruções fornecidas pelo protocolo do fabricante. Os meios YPD (Yeast Extract Peptone Dextrose), LB (Luria-Bertani) e de esporulação foram preparados de acordo com protocolos padrões (Ausubel *et al.* 1998). Para a seleção dos transformantes integrativos, quando necessário, o meio YPD foi complementado com o antibiótico geneticina, G418, a uma concentração de 200 µg/ml ou higromicina 300 µg/ml, após autoclavagem. Já o meio LB, foi complementado com ampicilina 100 µg/ml para seleção de transformantes. Para a preparação dos meios sólidos foram adicionados 2% de ágar. Técnicas de extração de DNA, transformação de leveduras pelo método acetato de lítio, esporulação, dissecação de tétrades e cruzamento de linhagens estão descritas em

protocolos padrões (Rose *et al.* 1990; Ausubel *et al.* 1998). Para esporulação, as leveduras foram plaqueadas em meios de esporulação e incubadas a 30°C por três dias. Após esporulação, as tétrades foram tratadas com a enzima Zymolyase. A dissecação de tétrades ocorreu através de micromanipulação. Após cruzamento de linhagens haploides, o *mating-type MATa/α* do diploide resultante foi confirmado por PCR, baseando-se na técnica descrita por (Huxley *et al.* 1990). Para crescimento, as linhagens de levedura foram incubadas a 30°C de um a três dias. O armazenamento das linhagens ocorreu em glicerol a -80°C. A Tabela 2 mostra os *primers* utilizados nesse trabalho.



Tabela 1 - Linhagens de levedura *Saccharomyces cerevisiae* utilizadas nesse estudo

Genótipos:				Descrições:	Referências:
<b>Diploides:</b>					
JAY270:	$MAT\alpha$ $MATa$			Isolado selvagem da linhagem industrial PE-2; não floculante	Argueso <i>et al.</i> 2009
JAY361:	$MAT\alpha$ $MATa$	$FLO8$ $flo8-1$	$FLO1$ $flo1\Delta$	Híbrido resultante do cruzamento de JAY291 e JAY479; não floculante	Argueso <i>et al.</i> 2009
ARY161:	$MAT\alpha$ $MATa$	$FLO8$ $FLO8::KanMX4$	$FLO1::HphMX4$ $flo1\Delta$	Híbrido resultante do cruzamento de JAY291 e ARY157; não floculante	Esse estudo
<b>Haploides:</b>					
JAY291:	$MATa$	$FLO8$	$flo1\Delta$	Esporo selvagem isolado de JAY270; não floculante	Argueso <i>et al.</i> 2009
JAY479:	$MAT\alpha$	$flo8-1$	$FLO1$	Linhagem de laboratório S288c selvagem (S97); petite ( $\rho 0$ ); não floculante	Steinmetz <i>et al.</i> 2002; Argueso <i>et al.</i> 2009
ARY153:	$MAT\alpha$	$flo8\Delta::HphMX4$	$FLO1$	Deleção de $FLO8$ com cassete $HphMX4-FLO8$ em JAY479; não floculante	Esse estudo
ARY156:	$MAT\alpha$	$FLO8::KanMX4$	$FLO1$	Reparo de $FLO8$ com inserção de pAR2 em ARY153; floculante	Esse estudo
ARY157:	$MAT\alpha$	$FLO8::KanMX4$	$FLO1::HphMX4$	Inserção de marca de resistência ligada ao $FLO1$ em ARY156; floculante	Esse estudo

Tabela 2- *Primers* oligonucleotídicos utilizados nesse estudo

Nome	Sequência 5' → 3' e objetivo
ARO72	TATAGGAATGTTGCATTT PCR-RFLP: região candidata localizada no Cromossomo VII, Forward
ARO73	ATTTCTGATAAATAACTT PCR-RFLP: região candidata localizada no Cromossomo VII, Reverse
ARO74	TTCCAGCCGCTAGACGAC PCR-RFLP: região candidata localizada no Cromossomo I, Forward
ARO75	ATCTTGGTGTTGGTGGCG PCR-RFLP: região candidata localizada no Cromossomo I, Reverse
ARO76	AGCAAATGGCAAACAACA PCR-RFLP: região candidata localizada no Cromossomo XV, Forward
ARO77	TAATGGCACTATTATATG PCR-RFLP: região candidata localizada no Cromossomo XV, Reverse
ARO78	TGGTGGTACTTCTTCATT PCR-RFLP: região candidata localizada no Cromossomo XII, Forward
ARO79	TAGATAGCCTCTGGAATA PCR-RFLP: região candidata localizada no Cromossomo XII, Reverse
ARO82	ATTGATAATCCTGATATG Homologia interna ao gene <i>KanMX4</i> , Reverse
ARO88	CTTGCCGCATGAACAGCCAAGCTTCTATATATTTGTTCTGCTCGAGCTCGTTTTTCGA <u>CA</u> Amplificação do cassete <i>HphMX4-FLO8</i> para deleção parcial de <i>FLO8</i> , Reverse
ARO89	ATTAAGGTTATGTGAGTTTCGTCCTGTAAAGTTGATCGGCGCTTGACGTGCGCAGCTC <u>AG</u> Amplificação do cassete <i>HphMX4-FLO8</i> para deleção parcial de <i>FLO8</i> , Forward
ARO90	TTAACATTGCGCATTTGAACC Homologia interna ao <i>FLO8</i> , confirmação da integração de <i>HphMX4-FLO8</i>

- ARO91 CCAATCTTGTATGTATGTGTC  
Amplificação parcial de *FLO8* de JAY291 para construção de pAR2, Forward
- ARO92 TTACCTTTGTGGGCTGTTTAG  
Amplificação parcial de *FLO8* de JAY291 para construção de pAR2, Reverse
- ARO95 CTATGGTATGTCCAATAAGAT  
Amplificação de *FLO1*, Forward
- ARO98 TCAGTCCATTCTAGAGTTGGCGATGGTTCATTAAT  
Amplificação de *FLO1*, sítio de restrição *XbaI* (sublinhado) adicionado à ponta 5', Reverse
- ARO101 TGGAAACCAGCTCTTCAC  
Região proximal ao gene *FLO1*, Forward
- ARO102 CATGAGGCTCTATAGAGA  
Região proximal ao gene *FLO1*, Reverse
- ARO115 TTTGAAAAAGTAGATATGTGTATCATAAAGAAAACGTTATTAAGCTGAAGCTTCGTAC  
GC  
Amplificação do cassete *HphMX4-FLO1*, homologia distal ao gene *FLO1*, Forward
- ARO116 GGTAATAATTCTTTTCATGGATTTTGTTCCTAAGACAATAAGGCCACTAGTGGATCT  
GA  
Amplificação do cassete *HphMX4-FLO1*, homologia distal ao gene *FLO1*, Reverse
- ARO117 TCTGCTGATATTTTAAATACTAGTGC  
Confirmação da integração do cassete *HphMX4-FLO1*, Reverse
- ARO118 CTTCAATTGCGTCCTTATTGCTGGC  
Confirmação da integração do cassete *HphMX4-FLO1*, Forward
- JAO14 AGGAGGGTATTCTGGGCCTCCATG  
Homologia interna aos plasmídeos pFA6aKanMX4, pAG32 e pAR2
- JAO15 ATGCGAAGTTAAGTGCGCAGAAAAG  
Homologia interna aos plasmídeos pFA6aKanMX4, pAG32 e pAR2
-

## Mapeamento genômico preliminar

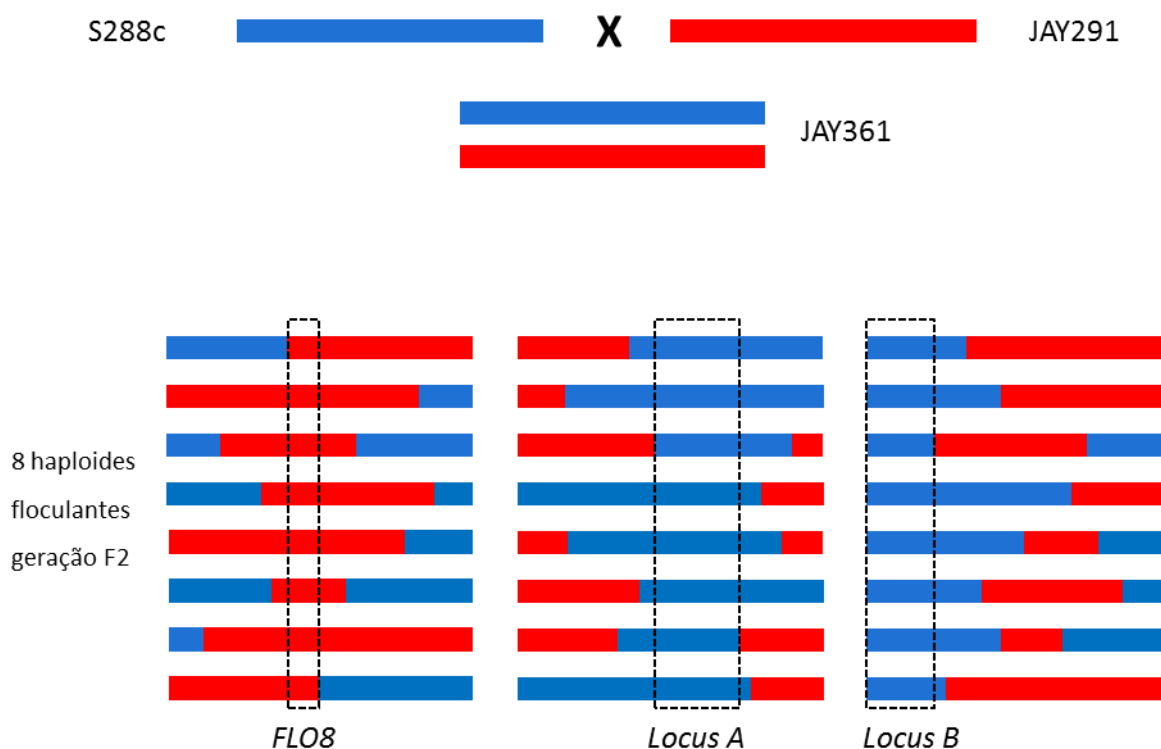
Primeiramente, realizamos um cruzamento entre a linhagem de laboratório S288c (JAY479) com a linhagem industrial PE-2 (JAY291), ambas não-floculantes. Nessa primeira análise, observamos os fenótipos de 120 indivíduos segregantes F2. Oito indivíduos floculantes selecionados entre essa população F2 tiveram seus genomas completos sequenciados, utilizando-se a plataforma “*Illumina HiSeq 2000*”, na “*University of North Carolina at Chapel Hill*”, Chapel Hill, Carolina do Norte, Estados Unidos. Os genomas dos oito indivíduos foram sequenciados em uma única calha (*lane*) de sequenciamento.

Analizamos as sequências geradas para os oito genomas através do programa “*CLC Genomics Workbench*”. Os “*reads*” gerados pelo sequenciamento foram mapeados no genoma da linhagem S288c, disponível no banco de dados “*Saccharomyces Genome Database*” (SGD), que possui alta qualidade de montagem e das sequências, os quais são sempre atualizados e disponíveis online. Também mapeamos as sequências de cada um dos oito indivíduos no genoma da linhagem JAY291 (Argueso *et al.* 2009). A cobertura média do sequenciamento foi de 77x.

Por conseguinte, interrogamos o genótipo de cada um dos oito indivíduos floculantes sequenciados em cada uma das posições de *Single Nucleotide Polymorphisms* (SNPs) existentes entre a linhagem de laboratório S288c e a linhagem industrial PE-2 (JAY291) (Figura 3). Foram analisados aproximadamente 40.000 SNPs segregantes entre os genomas dessas duas linhagens, o que significa aproximadamente um SNP a cada 300 pb. Nosso objetivo nessa etapa foi encontrar três regiões que co-segregavam com o fenótipo floculante dos indivíduos analisados (Figura 4).



**Figura 3 - Re-sequenciamento genômico e mapeamento dos *reads* no genoma referência de alta qualidade de S288c.** Os genomas dos esporos floculantes foram mapeados no genoma de S288c, depositado em SGD (representado pela seta azul) e então os genótipos dos haploides floculantes foram interrogados em 40.000 SNPs existentes entre S288c e JAY291. Os genomas dos indivíduos floculantes foram mapeados nos genomas de ambos os parentais, S288c e JAY291.

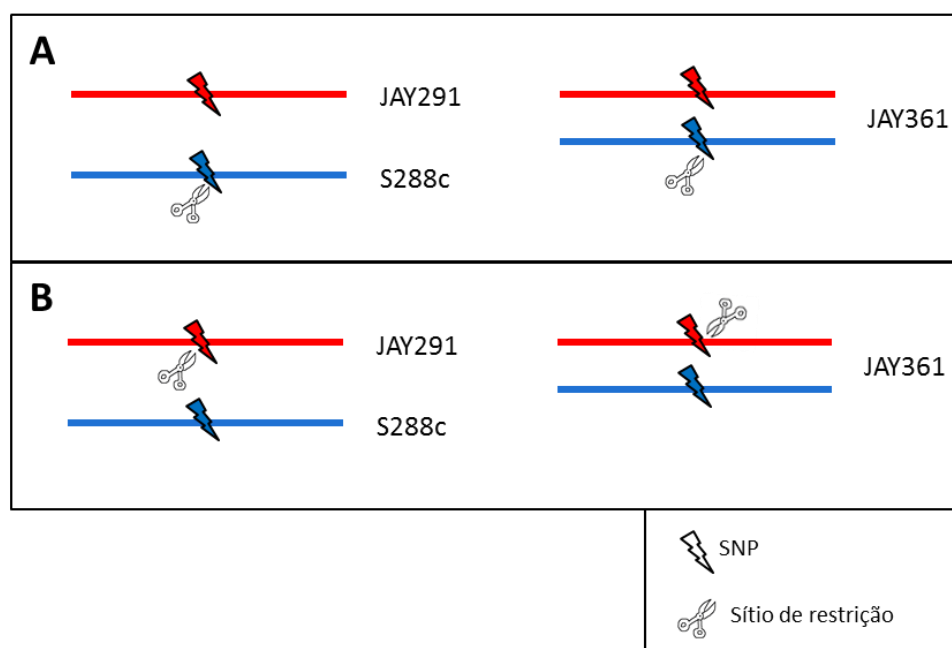


**Figura 4 - Sequenciamento e mapeamento da população flocculante F2.** Para identificar os genes que controlam o fenótipo flocculante observado na população F2 do cruzamento de S288c e JAY291, dentre os 120 haploides derivados de JAY361, selecionamos oito indivíduos e sequenciamos seus genomas completos, segregando os ~40.000 SNPs existentes entre os genomas parentais. Ao realizarmos o mapeamento, as sequências dos genomas dos flocculantes foram comparadas entre si e com os genomas parentais, com o objetivo de se encontrar regiões que co-segregam com o fenótipo flocculante dos oito indivíduos, podendo ser herdadas tanto de S288 (representado em azul) ou de JAY291 (representado em vermelho). Nossa expectativa nessa etapa é encontrar três genes não ligados que controlam o fenótipo não flocculante de JAY291. Uma dessas regiões é o segmento do Cromossomo V de JAY291, contendo alelo funcional de *FLO8* (Argueso *et al.* 2009).

### Refinando o mapeamento por PCR-RFLP (*Polymerase Chain Reaction - Restriction Fragment Length Polymorphism*)

Após encontrarmos cinco regiões candidatas no primeiro mapeamento genômico, foram analisados os fenótipos de 160 indivíduos adicionais, segregantes F2 do mesmo cruzamento de S288c (JAY479) e PE-2 (JAY291). Já sabíamos que a região localizada no Cromossomo V estava realmente associada com o fenótipo, analisada por PCR-RFLP nos 19 indivíduos flocculantes da primeira progênie F2 (Argueso *et al.* 2009). Dentre os novos

160 indivíduos F2, 16 flocculantes foram selecionados para genotipagem de outras quatro regiões candidatas também através da técnica PCR-RFLP. Para essa análise, selecionamos SNPs localizados nas regiões candidatas que criavam sítios de restrição específicos ou para a linhagem de laboratório (S288c) ou para a linhagem industrial (JAY291), como mostra a Figura 5.



**Figura 5 - Refinando o mapeamento por PCR-RFLP.** Para verificar o genótipo de dezesseis indivíduos flocculantes adicionais em quatro regiões candidatas, selecionamos SNPs que criam sítios de restrição específicos de S288c (A) ou de JAY291 (B).

Todos os *primers* foram desenhados para amplificação de fragmentos contendo aproximadamente 600 pb, sendo que *forward* e *reverse* distam, cada um, cerca de 200 e 400 pb do SNP/sítio de restrição selecionado, ou seja, quando houver o sítio de restrição, a digestão gera dois fragmentos, um de aproximadamente 200 pb e outro de aproximadamente 400 pb.

O PCR ocorreu pela reação a seguir, utilizando *MyTaq DNA Polymerase* (Bioline) e DNA genômico de cada um dos dezesseis indivíduos analisados: desnaturação inicial de

94°C por 5 minutos, seguida de 35 ciclos à 94°C por 30 segundos, temperatura de hibridização dos *primers* por 30 segundos, 72°C por 1min e extensão final à 72°C por 5 minutos.

Para análise da região candidata localizada no Cromossomo I, utilizamos os *primers* ARO74-75 para amplificação de PCR, seguido de reação de digestão pela enzima de restrição *SacII*. Já para a região candidata encontrada no Cromossomo VII, utilizamos os *primers* ARO72-73 e enzima de restrição *Sspl*. Para a região candidata encontrada no Cromossomo XII, ARO78-79 e enzima *Sspl* e para a localizada no Cromossomo XV, ARO76-77 e enzima *Hinfl*. Todas as enzimas de restrição utilizadas são da marca New England Biolabs Inc. (NEB).

O PCR para testarmos a amplificação ou não amplificação de *FLO1* no genoma das linhagens S288c e JAY291 foi realizado através de reação preparada da mesma maneira descrita acima, mas com os *primers* ARO101-102.

## **Construção do plasmídeo pAR2**

Com o intuito de se reparar o defeito *flo8-1* da linhagem de laboratório S288c construímos um plasmídeo, denominado como pAR2, contendo sequências parciais do alelo funcional *FLO8* de JAY291 e o gene de resistência a geneticina (*KanMX4*), que foi utilizado para integração por recombinação homóloga. O plasmídeo *pFA6aKanMX4* (Wach *et al.* 1994), que contém o marcador *KanMX4*, foi utilizado como base.

Inicialmente, nossa estratégia era construir um plasmídeo contendo o gene *FLO1* de S288c e o gene *FLO8* de JAY291 para inseri-lo em ambas as linhagens de laboratório e industrial. Esse plasmídeo chegou a ser construído, mas posteriormente decidimos mudar



a estratégia, apenas reparando o defeito *flo8* na linhagem de laboratório e ligando seu gene *FLO1* a um marcador (*HphMX4*), o que seria útil para selecionar linhagens F2 que herdassem o gene *FLO1* de S288c após seu cruzamento com JAY291. Houve a mudança de estratégia porque o gene *FLO1* possui muitas repetições internas que poderiam comprometer a fidelidade de amplificação e consequentemente sua funcionalidade.

Portanto, primeiro construímos um plasmídeo com os genes *FLO1*, *FLO8* e *KanMX4* e posteriormente o gene *FLO1* foi removido, resultando no plasmídeo pAR2 (Figura 6).

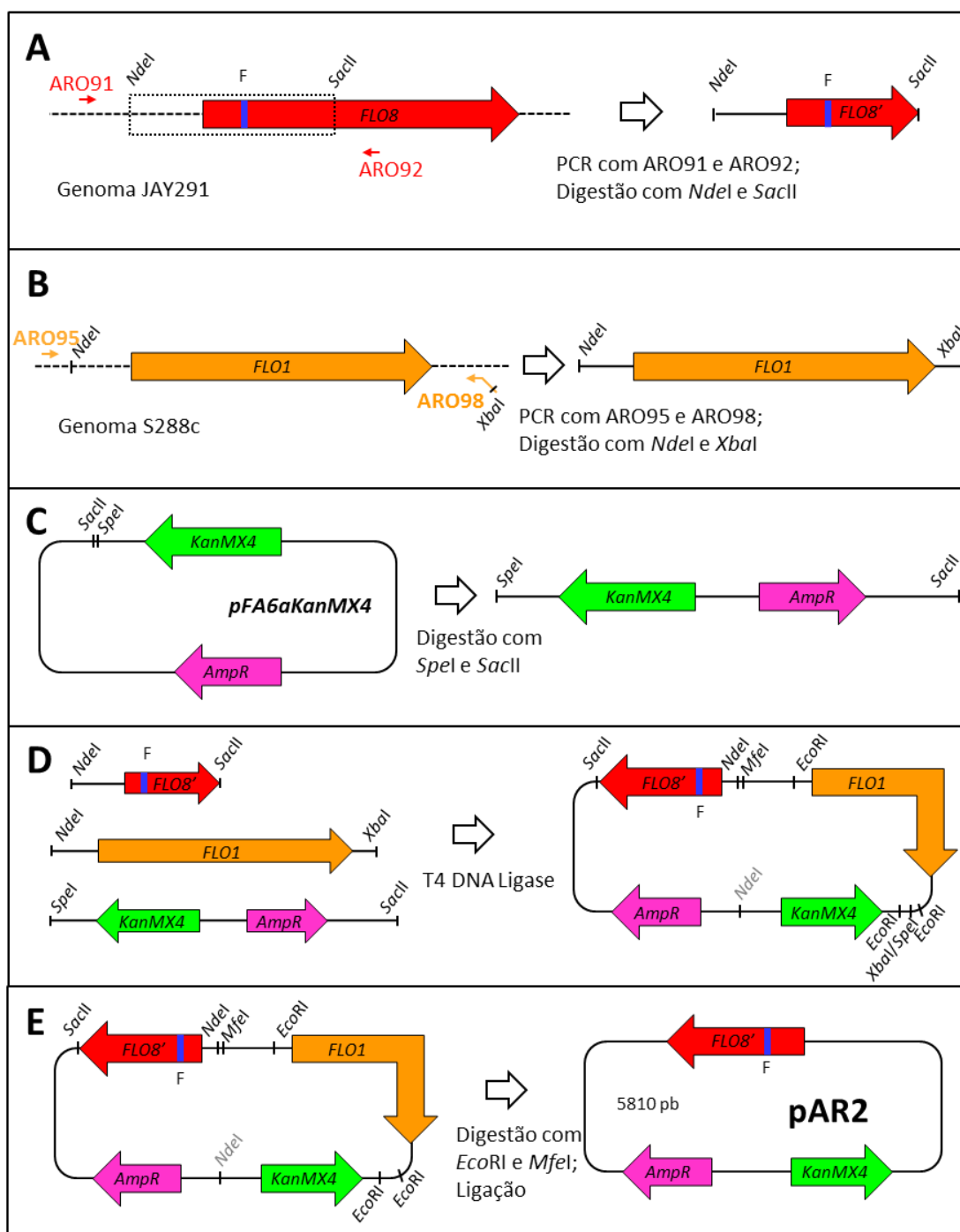
O gene *FLO1* foi amplificado do genoma de S288c (JAY479) com os *primers* ARO95 e ARO98 (contém uma cauda com o sítio de restrição *XbaI*) gerando um fragmento de 6,4 Kb. A reação de PCR foi gel purificada e após reação de digestão com as enzimas *XbaI* e *NdeI* resultou em um fragmento de 6,1 kb, gel purificado e reservado para inserção no plasmídeo juntamente com o fragmento de *FLO8*.

Sequências parciais do gene *FLO8* foram amplificadas com os *primers* ARO91 e ARO92 e DNA genômico da linhagem JAY291, utilizando *Velocity DNA Polymerase* (Bioline). A amplificação por PCR foi iniciada a 98°C por 50 seg e então realizada com 38 ciclos de 98°C por 10 seg, 55°C por 20 seg e 72°C por 2 min, com um passo final de 4 min a 72°C. O produto de amplificação de 2,3 Kb foi purificado de gel de agarose (*GeneJET Gel Extraction Kit*, Thermo Scientific) e subsequentemente digerido com as enzimas de restrição *SacII* e *NdeI* (New England Biolabs - NEB). O fragmento de interesse de 1,8 kb foi purificado de gel de agarose.

Por outro lado, o plasmídeo *pFA6aKanMX4* foi digerido com as enzimas *SacII* e *SpeI* (NEB) gerando dois fragmentos, sendo que o fragmento de 3921 pb foi isolado e purificado de gel de agarose e o fragmento de 20 pb descartado.

Os fragmentos de *FLO1*, *FLO8* e *pFA6aKanMX4* foram ligados utilizando-se a enzima *T4 DNA Ligase*, criando um plasmídeo de 11,9 kb, clonado por *E. coli* (Bioline) e extraído com “*GeneJET Plasmid Miniprep Kit*” (Thermo Scientific).

Finalmente, a remoção do gene *FLO1* foi realizada através da digestão do plasmídeo com as enzimas *MfeI* e *EcoRI* (NEB) e o fragmento de interesse de 5,8 kb, contendo sequência parcial do gene *FLO8* e o marcador *KanMX4*, foi purificado de gel de agarose e religado com *T4 DNA Ligase*, resultando no plasmídeo pAR2. Após a clonagem em *E. coli*, pAR2 foi isolado com utilização do “*GeneJET Plasmid Miniprep Kit*” (Thermo Scientific). Todas as etapas realizadas para construção de pAR2 estão representadas na Figura 6.



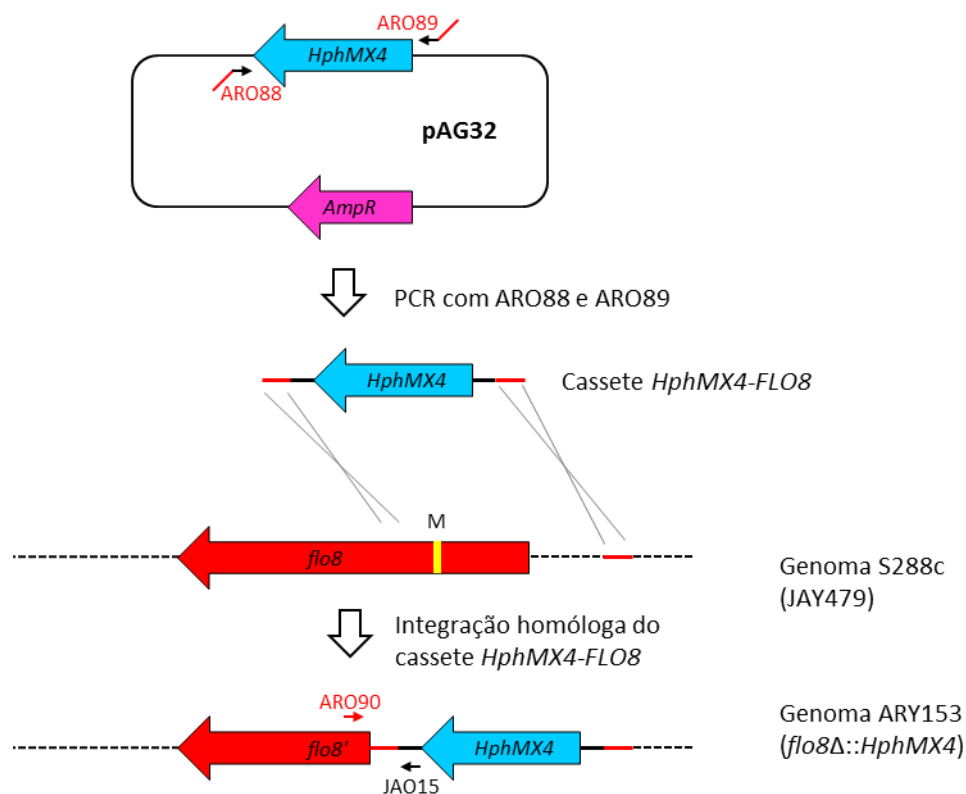
**Figura 6 - Representação das etapas realizadas para a construção do plasmídeo pAR2, utilizado para reparo do gene *FLO8* em S288c.** Em A, amplificação de seqüências parciais de *FLO8* do genoma da linhagem industrial JAY291, pelos primers ARO91 e ARO92, seguida de reação de digestão com enzimas de restrição *NdeI* e *SacII*. A amplificação de *FLO1* do genoma da linhagem de laboratório S288c, pelos primers ARO95 e ARO98, foi digerida com as enzimas de restrição *NdeI* e *XbaI* (B). Em C, digestão do plasmídeo *pFA6aKanMX4*, que contém o gene de resistência ao antibiótico geneticina, com enzimas *SpeI* e *SacII*. A etapa D mostra a ligação dos fragmentos gerados nas etapas anteriores (A-C) gerando um plasmídeo que contém os genes *FLO1*, *FLO8* e *KanMX4*. Para a construção de pAR2, ocorreu a remoção do gene *FLO1*, através de digestão com enzimas *EcoRI* e *MfeI*, do plasmídeo construído em (D), que foi religado, resultando em um plasmídeo final de 5810 pb, com seqüências parciais de *FLO8* de JAY291 e gene de resistência ao antibiótico geneticina (*KanMX4*) (E).

## **Reparo da mutação *nonsense flo8-1* e inserção do marcador *Hph* ligado ao gene *FLO1* no genoma de S288c**

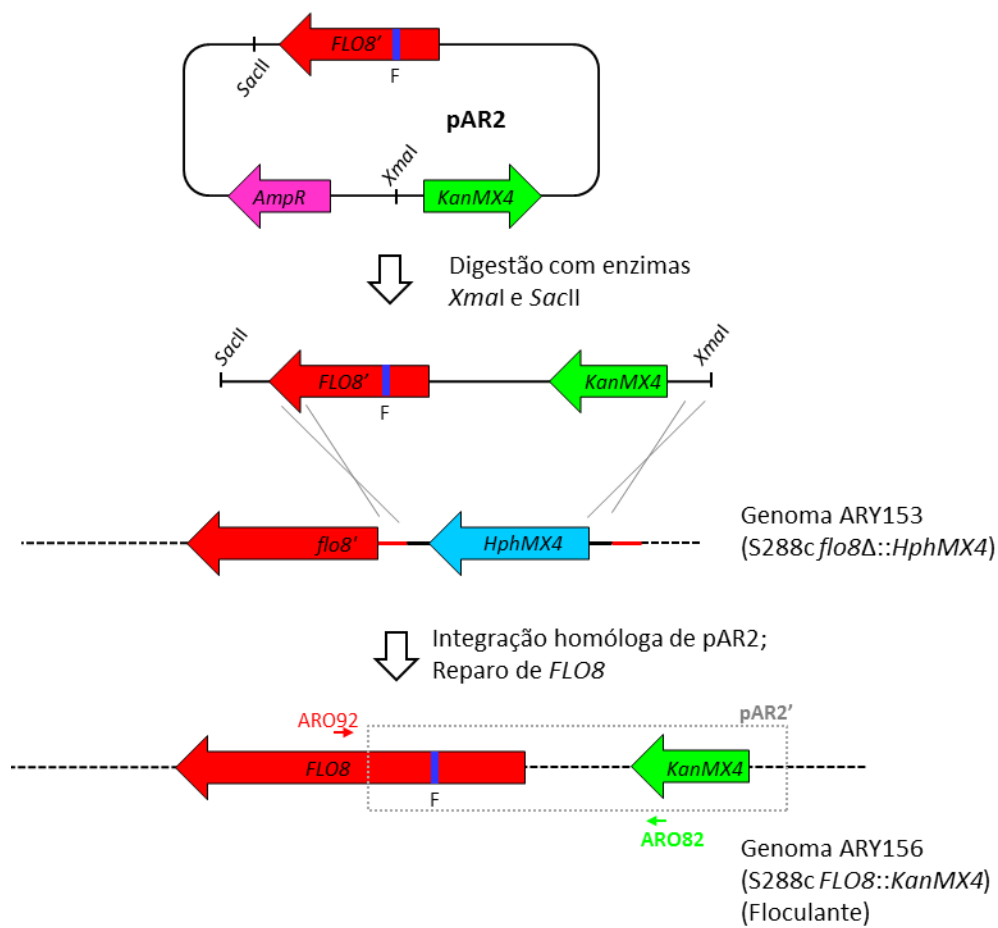
Primeiramente foi construído um cassete, nomeado de *HphMX4-FLO8*, para deleção parcial de *FLO8* no genoma da linhagem de laboratório S288c (JAY479). O cassete *HphMX4-FLO8* foi amplificado do plasmídeo pAG32 (Goldstein and McCusker 1999) com utilização da enzima “*MyTaq DNA Polymerase*” (Bioline) e os *primers* quiméricos ARO88-ARO89 (42 nucleotídeos homólogos com a região flanqueadora do locus alvo e 18 nucleotídeos para a amplificação *HphMX4*-pAG32). A reação ocorreu após desnaturação inicial de 5 min a 94°C, em 35 ciclos de 30 seg a 94°C, 30 seg a 55°C, 3 min a 72°C, com um passo final de 5 min a 72°C.

A integração correta de *HphMX4-FLO8* foi confirmada por PCR de colônia com os oligonucleotídeos ARO90 e JAO15 e a linhagem S288c com deleção de parte do gene *FLO8* foi salva como ARY153 (Figura 7).

Posteriormente, o plasmídeo pAR2 foi inserido no genoma de ARY153, removendo sequências do cassete *HphMX4-FLO8*, através de integração homóloga, reparando o defeito do gene *FLO8*. Antes da transformação em levedura, pAR2 foi digerido com as enzimas *SacII* e *XmaI*. A integração de pAR2 na posição genômica correta foi confirmada por PCR de colônia com os *primers* ARO82 e ARO92. A linhagem S288c com o *FLO8* reparado através da integração homóloga de pAR2 foi salva como ARY156 (Figura 8).



**Figura 7 - Deleção parcial de *flo8* em S288c (JAY479) pela integração homóloga do cassete *HphMX4-FLO8*.** O cassete *HphMX4-FLO8* foi construído através de amplificação a partir do plasmídeo pAG32, que possui gene de resistência ao antibiótico higromicina (*HphMX4*). Para a reação foram utilizados os *primers* quiméricos ARO88 e ARO89, possuindo 42 bases de homologia com o *locus* alvo de deleção (cauda vermelha dos *primers*) e 18 bases de homologia interna a pAG32 (parte representada em preto nos *primers*). Após o PCR, o cassete de 1,6 kb foi utilizado para transformação. A integração homóloga do cassete ocorreu no genoma de S288c (JAY479) e a nova linhagem, com deleção da região que possui mutação *nonsense flo8-1* (M), foi nomeada ARY153.

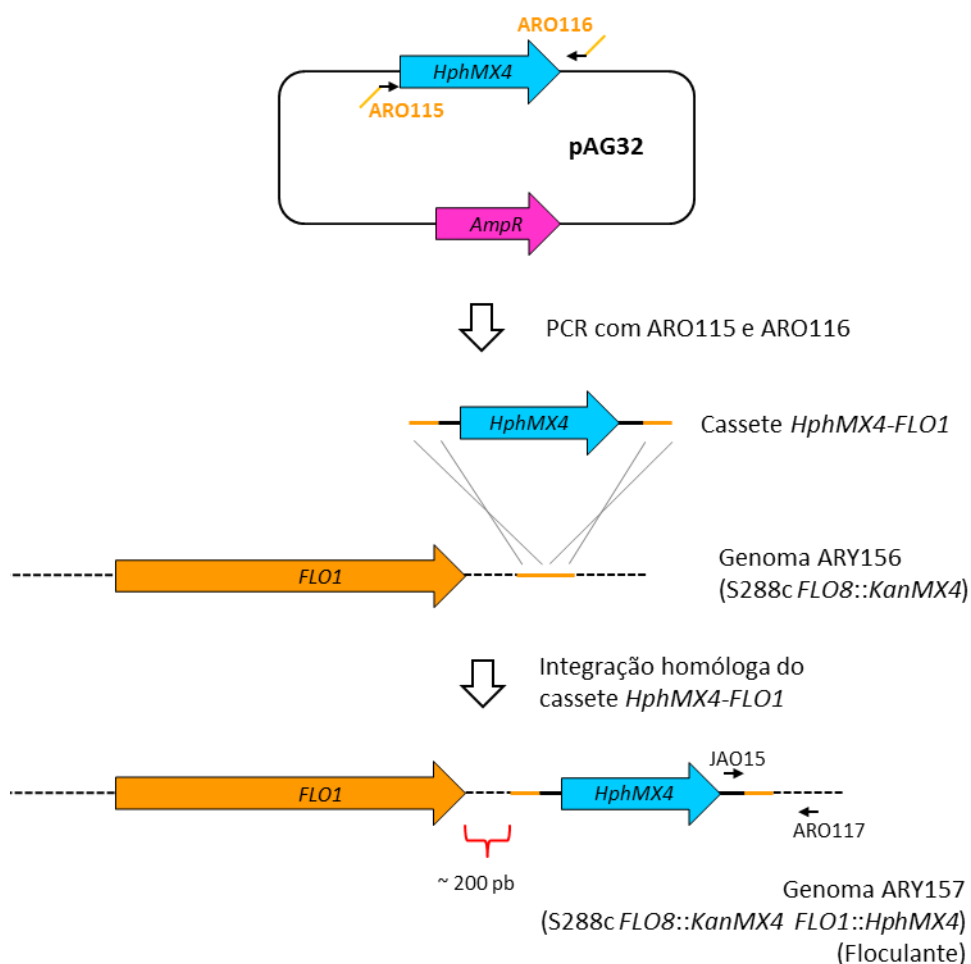


**Figura 8 - Reparo de *flo8* em S288c (ARY153) pela integração homóloga do plasmídeo pAR2.** O plasmídeo pAR2 possui sequências parciais do alelo funcional (F) de *FLO8* da linhagem industrial JAY291. Após digestão de pAR2 com as enzimas de restrição *XmaI* e *SacII*, ocorreu sua integração no genoma da linhagem ARY153, que possui deleção parcial de *flo8* pela integração do cassete *HphMX4*-*FLO8*. A região de homologia para integração de pAR2 ocorreu entre sequências de *FLO8* do plasmídeo e de ARY153 e também entre sequências de pAR2 e cassete *HphMX4*-*FLO8*. A inserção de pAR2 foi confirmada por PCR, utilizando-se um *primer* externo à integração e interno ao gene *FLO8* (ARO92) e um *primer* interno ao gene *KanMX4*, específico de pAR2 (ARO82). A confirmação da correta integração e consequente reparo do gene *FLO8* também puderam ser observados pela mudança de fenótipo da linhagem, que passa a ser floculante e foi nomeada ARY156.

## Inserção do marcador *Hph* ligado ao gene *FLO1* no genoma de S288c

O cassete *HphMX4-FLO1*, desenvolvido para ligar o gene *FLO1* do genoma de S288c a um marcador de resistência ao antibiótico higromicina, foi amplificado de pAG32 com o mesmo programa utilizado para amplificação do cassete *HphMX4-FLO8*, porém com os oligonucleotídeos quiméricos ARO115-ARO116. O cassete *HphMX4-FLO1* foi inserido a aproximadamente a 200 pb (180 pb) distal ao gene *FLO1*, no genoma da linhagem ARY156. A linhagem S288c (*FLO8::KanMX4 FLO1::HphMX4*) foi salva como ARY157 (Figura 9).

A inserção foi confirmada por PCR com os *primers* ARO117-JAO15 após a seleção dos transformantes em placas contendo os antibióticos geneticina e higromicina. O produto amplificado pelo PCR digerido com a enzima *TaqI* para checar se o marcador foi inserido no local correto (detalhado em resultados). Após pré-seleção do candidato pelo padrão de restrição, os fragmentos amplificados através de duas reações PCR, uma realizada com os *primers* ARO117-JAO15 e outra com os *primers* ARO116-JAO14, foram sequenciados. O sequenciamento ocorreu em um equipamento “ABI 3130xL Genetic Analyzer” (Applied Biosystems), em “Proteomics and Metabolomics Facility”, localizada na “Colorado State University”, Fort Collins, Colorado, EUA. As sequências foram analisadas através do software “CLC Genomics Workbench”.



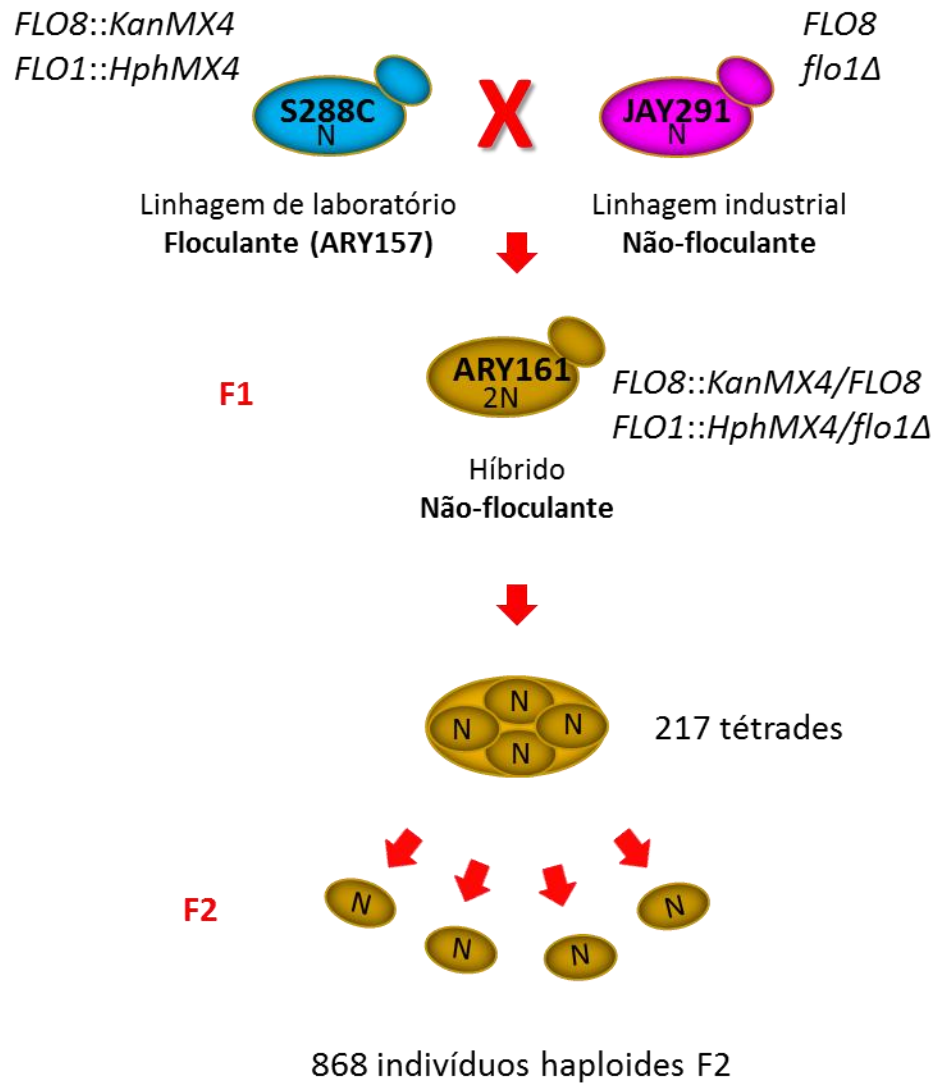
**Figura 9 - Inserção do cassete *HphMX4-FLO1* em uma região ~200 pb distal ao *FLO1* de S288c (ARY156) por integração homóloga.** O cassete *HphMX4-FLO1* foi construído através de amplificação a partir do plasmídeo pAG32, que possui gene de resistência ao antibiótico higromicina (*HphMX4*). Para a reação foram utilizados os *primers* quiméricos ARO115 e ARO116, possuindo 42 bases de homologia com o *locus* alvo, distal a *FLO1* (cauda laranja dos *primers*) e 18 bases de homologia interna a pAG32 (parte representada em preto nos *primers*). Após o PCR, o cassete de 1,8 kb foi utilizado para transformação. A integração homóloga do cassete ocorreu no genoma de ARY156 e a nova linhagem, agora com marca de seleção ligada ao gene *FLO1* passou a ser chamada de ARY157.



## **Mapeamento genômico em “bulks” - *Bulk Segregant Analysis* (BSA) aliada a sequenciamento de última geração - *Next Generation Sequencing* (NGS)**

Para um segundo mapeamento, realizamos uma nova estratégia, diferente da utilizada no primeiro caso. Nesse caso, primeiramente realizamos um cruzamento entre a linhagem de laboratório S288c (*FLO8::KanMX4 FLO1::HphMX4*), nomeada como ARY157, com a linhagem industrial JAY291 (*FLO8 flo1Δ*).

Selecionamos 217 tétrades completas do híbrido resultante (ARY161), ou seja, que possuíam os quatro haploides da tétrade viáveis e que também possuíam segregação 2:2 tanto para resistência à higromicina, quanto para resistência à geneticina (dois esporos cresciam em meio contendo geneticina e dois esporos cresciam em meio contendo higromicina), totalizando 868 indivíduos segregantes haploides F2. A observação do crescimento nas placas YPD contendo higromicina e YPD contendo geneticina foi realizada após três dias em estufa a 30°C (Figura 10).



**Figura 10 – Nova população para mapeamento.** Primeiramente foi realizado um cruzamento entre a linhagem S288c/ARY157 ( $FLO8::KanMX4$   $FLO1::HphMX4$ ) e JAY291 ( $FLO8$   $flo1\Delta$ ), gerando o híbrido ARY161. 217 tétrades completas de ARY161 foram selecionadas e 868 indivíduos haploides F2 analisados fenotipicamente. Somente linhagens F2 portadoras do marcador  $HphMX4-FLO1$  foram selecionadas para o novo sequenciamento e mapeamento.

O próximo passo foi analisar o fenótipo de cada um dos 868 indivíduos e classificá-los como floculante ou não floculante. Com esse intuito, amostras de cada indivíduo foram inoculadas separadamente em tubos contendo 5 ml de meio líquido YPD. O tipo de crescimento em meio líquido (floculante/não floculante) foi observado e classificado após dois dias de crescimento em estufa a 30°C e rotação a 120 rpm. Essa classificação ocorreu a partir de observação das células a olho nu, logo após vortexar os tubos por um tempo aproximado entre 5 e 10 segundos. A classificação ocorreu por comparação das linhagens quanto ao tempo de sedimentação após a retirada do tubo do vórtex. Para essa classificação, as linhagens foram divididas em três grupos diferentes: não floculantes, floculantes e as mais floculantes, que sedimentavam logo após a retirada do tubo do vórtex. Para as linhagens floculantes, não somente sedimentavam mais rapidamente do que as não floculantes, como também foi observada a formação de flocos, típicos do fenômeno e que não ocorre em linhagens não floculantes.

Para o sequenciamento e mapeamento, somente os indivíduos portadores do marcador *HphMX4-FLO1* foram selecionados e então sequenciados em “bulks” e não individualmente como ocorreu no mapeamento preliminar realizado nesse trabalho. Para a realização do BSA-Seq, realizamos uma nova classificação dos indivíduos, dividindo-os em dois grupos para sequenciamento.

O primeiro grupo continha somente os indivíduos não floculantes e resistentes à higromicina (Grupo 1). Já o segundo grupo, possuía os indivíduos mais floculantes, ou floculantes extremos, e também resistentes à higromicina (Grupo 2).

Cada segregante cresceu individualmente YPD líquido por três dias a 30°C e 120 rpm. Foi então coletada a mesma quantidade de cultura de cada indivíduo selecionado, imediatamente após vortexar e colocado em um *Erlenmeyer* para mistura das culturas dos indivíduos de cada grupo. Foram utilizados dois recipientes *Erlenmeyer*, um para cada grupo de sequenciamento, cada um contendo uma barra magnética para mistura. Cada *Erlenmeyer* foi mantido em gelo até completarmos o inóculo de todas as amostras pertencentes ao grupo. Após a agitação com barra magnética por 15 minutos, amostras do conteúdo total foram transferidas para tubos de 15 ml, aos quais foram centrifugados, os sobrenadantes descartados e os pellets foram utilizados para a extração de DNA.

O sequenciamento ocorreu em “*Illumina HiSeq 2000*” (2x 100pb), na “*University of North Carolina at Chapel Hill*”, Chapel Hill, Carolina do Norte, Estados Unidos, e cada amostra de DNA, representando um grupo (pool) de indivíduos segregantes F2 selecionados, foi sequenciada individualmente em uma calha (*lane*), com o intuito de se obter a máxima qualidade dos dados em cada grupo.

As sequências geradas foram analisadas pelo programa “*CLC Genomics Workbench*” e os “*reads*” gerados pelo sequenciamento foram mapeados no genoma da linhagem de laboratório S288c, disponível no banco de dados “SGD”.

As frequências alélicas, ou as porcentagens de alelos herdados de cada parental, S288c e JAY291, foram calculadas em cada posição dos 40.000 SNPs existentes entre a linhagem de laboratório S288c e a industrial JAY291.

As estatísticas para validarmos os dados gerados pelo mapeamento em *BSA-Next Generation Sequencing* e consequentemente, os loci associados com o fenótipo, foram

calculadas pelo professor Dr. Paul Magwene, da “*Duke University*”, localizada em Durham, Carolina do Norte, Estados Unidos, conforme metodologia descrita em seu artigo (Magwene *et al.* 2011). Foram calculadas as estatísticas para os dados gerados em cada “bulk” individualmente e também do “bulk” não flocculante (Grupo 1) versus o “bulk” flocculante (Grupo 2). Para analisarmos os loci associados e posteriormente genes candidatos e relacionados ao fenótipo, utilizamos os dados de comparação estatística de um grupo versus o outro: Grupo 1 (não flocculantes e resistentes a higromicina) versus Grupo 2 (flocculantes e resistentes a higromicina).

## RESULTADOS

### Mapeamento genômico preliminar

Dadas às primeiras observações descritas na introdução desse capítulo, realizamos um mapeamento preliminar com o intuito de localizarmos regiões que co-segregam com o fenótipo de flocculação e controlam o fenótipo não flocculante de JAY291. Para isso, primeiramente selecionamos oito entre os dezenove indivíduos F2 flocculantes e realizamos o sequenciamento completo de seus genomas a uma cobertura relativamente baixa (77x). Os genomas dos oito indivíduos foram mapeados no genoma referência da linhagem de laboratório S288c e seus genótipos foram interrogados em aproximadamente 40.000 posições de SNPs existentes entre a linhagem de laboratório S288c e a linhagem industrial JAY291.

Pela previsão da taxa fenotípica, nossa expectativa era de encontrarmos três regiões co-segregando com o fenótipo flocculante dos oito indivíduos analisados, ou seja,

regiões que estavam presentes nos genomas de todos os indivíduos floclantes, sendo herdadas de S288c ou JAY291 (Figura 4).

Através da comparação entre os genótipos dos oito indivíduos encontramos cinco regiões candidatas que co-segregavam com o fenótipo de floclação. A primeira região estava localizada no Cromossomo I, com 8,2 kb (185626-193823 pb); a segunda, localizada no Cromossomo V, possuía 112 kb (342715-454722 pb); a terceira estava localizada no Cromossomo VII, com 9,8 kb (971314-981052 pb); a quarta no Cromossomo XII, com 1,8 kb (981506-983258 pb) e a quinta região candidata, estava localizada no Cromossomo XV, contendo 10,5 kb (226291-236809 pb). Os números apresentados entre parênteses correspondem às posições dos marcadores SNPs no genoma da linhagem JAY291.

Analisando essas regiões no genoma da linhagem S288c (SGD), a região candidata localizada no Cromossomo I está na ponta direita a partir do gene *SWH1* e o último marcador está em *FLO1*. Já a região candidata do Cromossomo V, corresponde à região que vai do gene *TSC11* ao gene *LSM5*. Para a região candidata localizada no Cromossomo VII, temos os genes de *TNA1* a *MES1*. Já para o Cromossomo XII, temos uma região de ponta direita a partir do gene *NBP1*, com último marcador em *GAB1*. Finalmente, a região candidata localizada no Cromossomo XV, compreende a região que vai do gene *SPE2* ao gene *YOL047C*.

O número de regiões candidatas encontradas, acima do esperado, poderia ser devido ao baixo número de indivíduos analisados e, por isso, estavam co-segregando com a população F2 apenas coincidentemente, sem estarem de fato relacionadas com o

fenótipo. Por esse motivo, resolvemos aumentar o número de segregantes analisados, aumentando consequentemente a população F2 e observando os fenótipos dos novos segregantes. Dessa vez, selecionamos mais quarenta tétrades completas do híbrido JAY361, resultando em 160 indivíduos F2 segregantes adicionais para análise. Dos 160 novos indivíduos analisados, 16 eram floculantes e 144 não floculantes, novamente uma proporção fenotípica de 1:8.

Resumindo, primeiramente analisamos 120 indivíduos F2, onde 19 eram floculantes. Em uma segunda análise, de 160 haploides segregantes, 16 eram floculantes. Somando-se o número total de indivíduos analisados, temos uma população F2 de 280 haploides, sendo 35 floculantes. No número total também temos novamente, a proporção fenotípica de 1:8 floculantes para não floculantes, esperada para uma característica controlada por três genes não ligados.

### **Genotipagem adicional por PCR-RFLP: refinando o mapeamento**

Para uma segunda análise, selecionamos os 16 indivíduos floculantes, da população F2 adicional de 160 indivíduos, para genotiparmos as cinco regiões candidatas encontradas no mapeamento preliminar, para validar regiões verdadeiramente associadas ao fenótipo.

Essa genotipagem dessa vez não ocorreu pelo sequenciamento completo dos genomas dos segregantes, mas sim através da técnica PCR-RFLP. Assim, selecionamos SNPs localizados nas regiões candidatas que criavam sítios de restrição específicos para somente em um dos parentais, ou seja, específicos ou para a linhagem de laboratório S288c ou para a linhagem industrial JAY291 (Figura 5). Dessa maneira, após a realização de

PCR naquela região, seguido de reação de digestão com enzimas de restrição, teríamos um parental com digestão do fragmento amplificado, um parental onde a digestão não ocorria e o híbrido, com os dois padrões de banda no gel, apresentando um fragmento digerido e um fragmento não digerido.

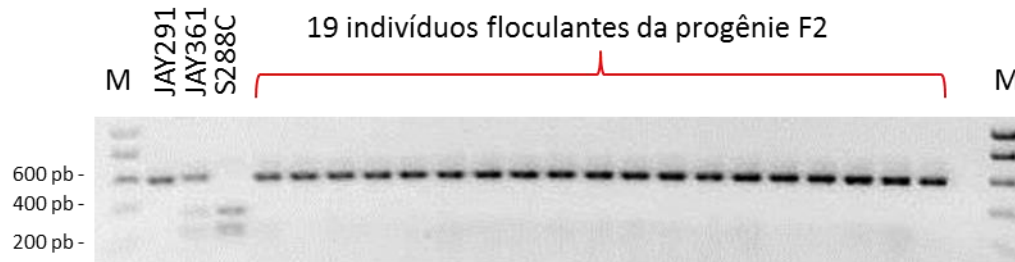
Para a região candidata localizada no Cromossomo I, a enzima de restrição *SacII* possui digestão específica para S288c. Da mesma forma, para a região candidata encontrada no Cromossomo VII, a enzima de restrição *SspI* também possui digestão específica para S288c. Já para a região candidata encontrada no Cromossomo XII, a enzima *SspI* tem digestão específica para JAY291 e finalmente, para a região localizada no Cromossomo XV, a enzima *HinfI* possui digestão específica para S288c.

Conforme esperado, quando havia o sítio de restrição presente no fragmento de aproximadamente 600 pb amplificado do genoma dos parentais, híbrido e floculantes analisados, a digestão resultava em dois fragmentos, um de aproximadamente 200 pb e outro de aproximadamente 400 pb. O híbrido apresentou as três bandas, correspondentes aos três fragmentos, um de 600 pb, não digerido, herdado do parental que não possui o sítio de restrição e os outros dois, de 200 e 400 pb, resultantes da digestão do fragmento herdado do parental que possui o sítio para a enzima.

Os resultados mostraram que duas das cinco regiões candidatas estavam associadas com o fenótipo floculante. Já sabíamos que uma dessas regiões correspondia ao gene *FLO8*, regulador global do fenômeno de floculação, pela co-segregação do alelo funcional de JAY291 em toda a primeira progênie F2 floculante de 19 indivíduos (Figura 11), mostrando o requisito de um *Flo8p* funcional para a floculação (Argueso *et al.* 2009).

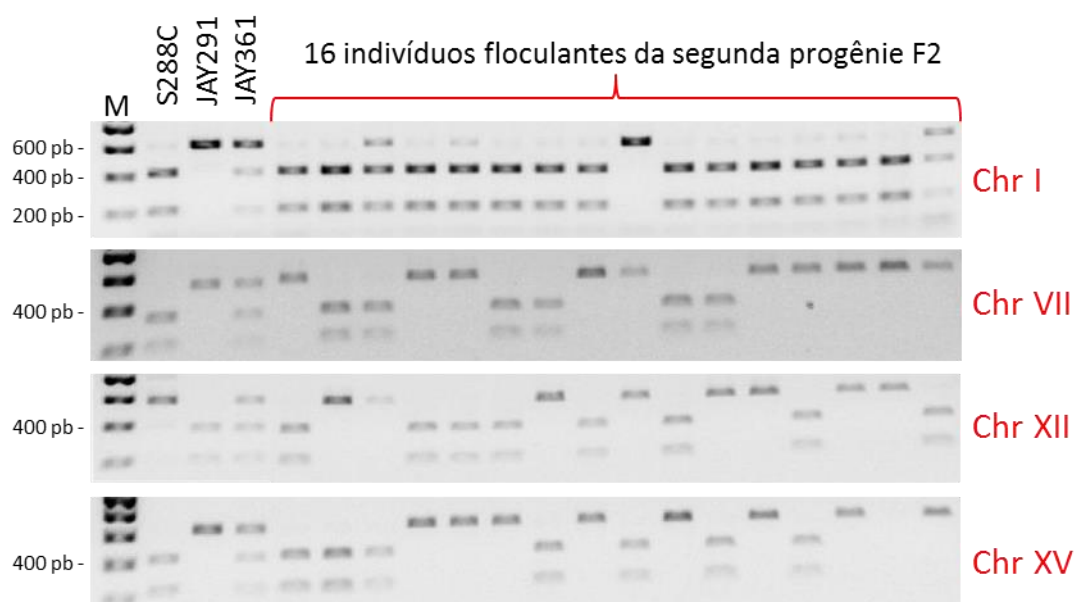


O gene *FLO8* está localizado na região do Cromossomo V encontrada no mapeamento genômico.



**Figura 11 - Co-segregação do alelo funcional *FLO8* (JAY291) na progênie F2 floculante.** A linhagem S288c possui uma mutação *nonsense* no gene *FLO8* (*flo8-1*), ligada a outro SNP que cria um sítio de restrição específico para a enzima *Sall*. Essa enzima não digere o fragmento de JAY291 (*FLO8*). Todos os dezenove haploides floculantes herdaram o alelo *FLO8* funcional de JAY291. Marcador de peso molecular “Hyperladder I” (1 kb) Bioline (M).

Das outras quatro regiões candidatas, analisadas na progênie F2 adicional de dezesseis indivíduos, uma delas co-segregava com o fenótipo floculante dos segregantes, a localizada no Cromossomo I. Quinze dos dezesseis indivíduos floculantes apresentaram o mesmo padrão de banda do parental S288c, indicando que todos possuem essa região herdada da linhagem de laboratório. Apenas um indivíduo apresentou o padrão da linhagem de laboratório JAY291, indicando provavelmente a ocorrência de um *crossing-over* entre a região associada e o marcador (SNP) escolhido (Figura 12).



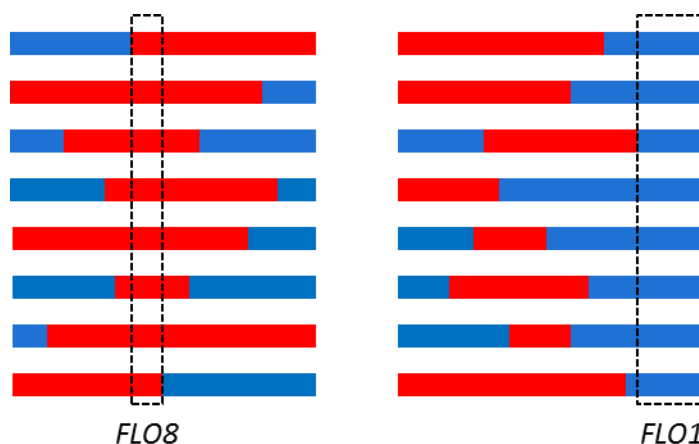
**Figura 12 - Genotipagem de quatro regiões candidatas em dezesseis indivíduos F2 floculantes adicionais por PCR-RFLP.** Géis de agarose com amostras de reação de PCR, após digestão com enzimas de restrição. A região localizada no Cromossomo I está ligada ao fenótipo, com co-segregação do alelo *FLO1* de S288c na progênie floculante. As outras regiões, localizadas nos Cromossomos VII, XII e XV não estão ligadas ao fenótipo floculante dos indivíduos. Marcador de peso molecular “Hyperladder I” (1 kb) Bioline (M).

A análise das outras três regiões candidatas nos dezesseis segregantes floculantes mostrou que não estavam associadas com o fenótipo. A região candidata analisada no Cromossomo XII mostra exatamente a frequência esperada para regiões não associadas, 50% ou oito indivíduos possuem a região herdada de S288c e os outros oito indivíduos possuem a região herdada de JAY291. Os padrões de herança apresentados para as regiões localizadas nos Cromossomos VII e XV também mostraram que essas duas não estão associadas com o fenótipo floculante dos dezesseis indivíduos analisados (Figura 12).

A região que co-segrega com o fenótipo floculante corresponde ao final do braço direito do Cromossomo I e analisando-a no genoma da linhagem S288c, através das sequências depositadas no programa SGD, vimos que é onde está presente o gene *FLO1*,

um dos principais genes já descritos que controla a característica floculante de linhagens de levedura.

Portanto, os resultados gerados pelo primeiro mapeamento mostraram que todos os segregantes floculantes possuíam *FLO8* herdado da linhagem industrial JAY291 e o *FLO1* herdado linhagem de laboratório S288c (Figura 13).

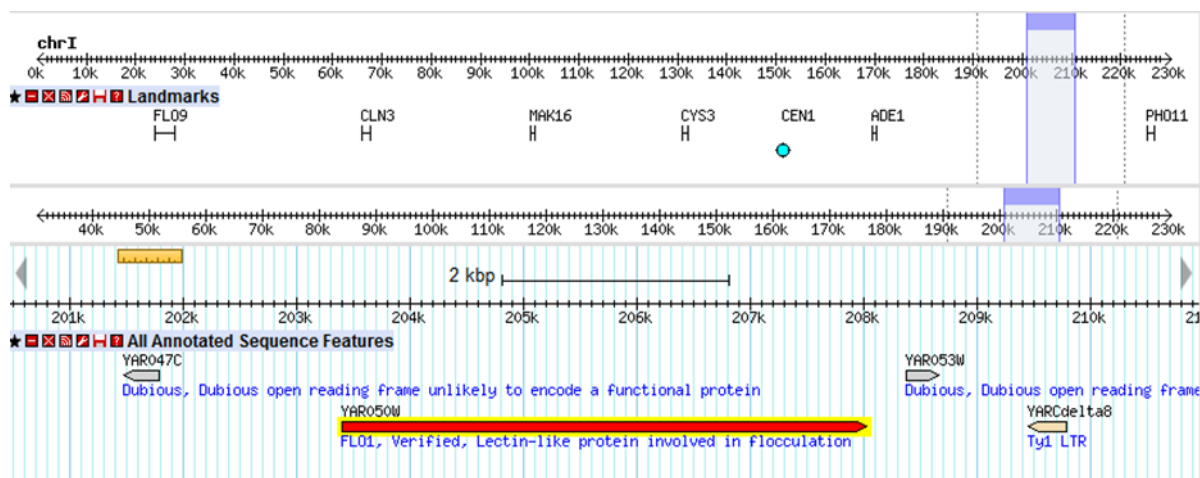


**Figura 13 - Regiões associadas com floculação.** Refinando o mapeamento PCR-RFLP, validamos duas dentre as cinco regiões candidatas relacionadas com o fenótipo floculante dos indivíduos. Todos os dezesseis esporos da segunda progênie F2 herdaram o alelo *FLO8* de JAY291 e o alelo *FLO1* de S288c, mostrando que essas regiões estão verdadeiramente associadas à floculação. As outras três regiões candidatas encontradas nos dados de sequenciamento foram apenas coincidência e não estão associadas com o fenótipo.

### ***FLO1* nos genomas de S888c e JAY291**

Vimos que o fenótipo floculante dependia da presença do alelo funcional Flo8p da linhagem industrial JAY291. A linhagem de laboratório S288c possui uma mutação *nonsense flo8-1*, que o torna inativo (Liu *et al.* 1996).

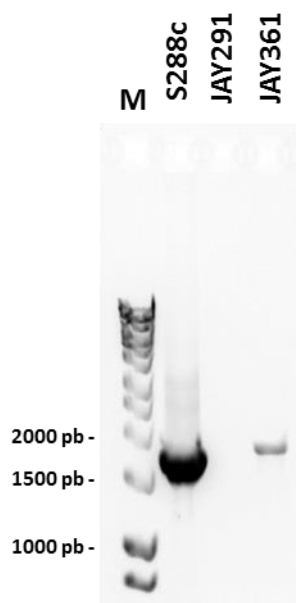
Também foi observado que o fenótipo era dependente da presença do alelo *FLO1* da linhagem de laboratório. O gene *FLO1* está localizado em uma região subtelomérica do Cromossomo I de S288c (Figura 14).



**Figura 14 – Gene *FLO1* no genoma de S288c.** No genoma de S288c, disponível em SGD (*Saccharomyces Genome Database*), o gene *FLO1* está localizado em uma região subtelomérica, na ponta direita do Cromossomo I.

Com essa observação, o próximo passo foi analisar o gene *FLO1* no genoma da linhagem industrial PE-2. Realizamos uma busca por esse alelo no genoma de JAY291 e surpreendentemente, observamos que ele não está presente nessa linhagem, explicando o resultado obtido pelo mapeamento genômico e genotipagem por PCR-RFLP.

Após avaliarmos as sequências genômicas de JAY291 e constataremos que esta não apresentava o alelo *FLO1*, testamos ainda por PCR a amplificação ou não amplificação de uma região localizada a aproximadamente 2790 pb proximal ao *FLO1*, que sabíamos que a linhagem JAY291 já não possuía em seu genoma. O resultado do PCR ocorreu conforme esperado, com amplificação da região (~1793 pb) para S288c e híbrido JAY361 e não amplificação para JAY291 (Figura 15).



**Figura 15 – Gene *FLO1* no genoma de JAY291.** Após constatararmos que o gene *FLO1* não está presente no genoma de JAY291 pela análise de suas sequências, testamos por PCR a amplificação de uma região distal ao *FLO1* no genoma de S228c, a não amplificação no genoma de JAY291 e amplificação para o híbrido entre as duas linhagens JAY361. A região amplificada tem aproximadamente 1793 pb. Marcador de peso molecular “Hyperladder I” (1 kb) Bioline (M).

Os resultados mostraram que a presença de alelos funcionais dos genes *FLO1* e *FLO8* é necessária para que a floculação ocorra. Porém, o híbrido JAY361, resultante do cruzamento de S288c e JAY291 possui esses alelos e não flocula, indicando que há pelo menos mais um gene controlando o fenótipo. Esse poderia ser o terceiro gene previsto pela proporção fenotípica inicialmente.

### Mapeamento genômico em “bulks”

No mapeamento realizado preliminarmente, esperávamos encontrar três genes não ligados associados com floculação. Porém, somente dois genes foram de fato encontrados, *FLO1* e *FLO8*. Vimos também que esses dois genes são necessários para que a floculação ocorra, mas não suficientes, indicando que há pelo menos mais um gene envolvido no processo.

Construímos então, uma nova estratégia de mapeamento, onde todos os segregantes a serem analisados teriam presentes em seus genomas o gene *FLO8* de JAY291 e o gene *FLO1* de S288c, facilitando a identificação de outras regiões também associadas ao fenótipo. Com esse objetivo, construímos uma linhagem S288c com *FLO8* funcional e *FLO1* marcado. Assim, realizando um novo cruzamento da linhagem S288c construída com a linhagem JAY291 (*FLO8 flo1Δ*) todos os segregantes teriam o *FLO8* funcional e poderíamos selecionar àqueles com o gene *FLO1* herdado de S288c, através da marca de seleção. Também aumentamos significativamente o número de indivíduos analisados e isso pôde ser realizado através do sequenciamento e mapeamento genômico em “bulks”.

### **Construção do plasmídeo pAR2**

Para reparo do alelo mutado *flo8-1* em S288c, construímos um plasmídeo, designado pAR2, contendo sequências parciais do alelo funcional *FLO8* de JAY291 e marca de seleção, de resistência ao antibiótico geneticina (*KanMX4*). O plasmídeo final obtido tem 5810 pb. Sua construção está descrita em “materiais e métodos” desse capítulo e ilustrada na Figura 6.

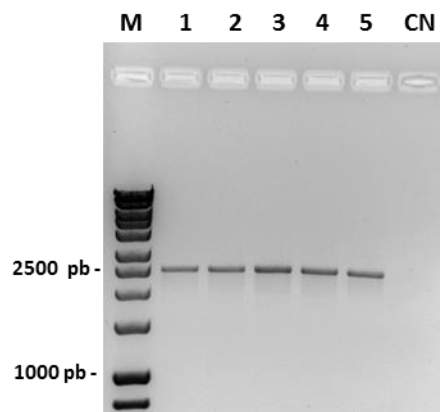
### **Construção de uma linhagem S288c *FLO8***

Primeiramente, foram deletadas sequências parciais do gene *FLO8* na linhagem de laboratório S288c (JAY479) através da recombinação homóloga do cassete *HphMX4-FLO8*. Esse cassete contém o gene de resistência a higromicina (*HphMX4*), homologia de 42 bases de cada lado flanqueando o *locus* alvo e foi amplificado como um fragmento de 1,6 kb a partir do plasmídeo pAG32 (Goldstein and McCusker 1999).

A confirmação da inserção do cassete *HphMX4-FLO8* ocorreu por PCR, após seleção em meio contendo o antibiótico higromicina, com um *primer* interno ao cassete (JAO15) e um *primer* externo à integração (ARO90), como mostra a Figura 7. O fragmento esperado de aproximadamente 280 pb foi amplificado, confirmando a integração. A linhagem obtida, com deleção parcial de *FLO8* pela inserção do cassete *HphMX4-FLO8* foi denominada ARY153.

A seguir, o fragmento de 3,3 kb gerado após digestão do plasmídeo pAR2 com as enzimas de restrição *SacII* e *XmaI* foi inserido no genoma da linhagem ARY153. A homologia para integração de pAR2 ocorreu entre sequências do gene *FLO8* da linhagem ARY153 e do plasmídeo pAR2 e entre sequências do plasmídeo e o cassete *HphMX4-FLO8*, como mostra a Figura 8. Portanto, a integração de pAR2 remove sequências do cassete *HphMX4-FLO8* e consequentemente a linhagem transformada passa a ser sensível à higromicina (gene de resistência estava presente no cassete *HphMX4-FLO8*) e resistente à geneticina (gene de resistência presente no plasmídeo pAR2).

Após seleção em meios contendo os antibióticos higromicina e geneticina, a integração de pAR2 foi confirmada por PCR, com amplificação pelos *primers* ARO82 (homologia interna ao gene *KanMX4* de pAR2) e ARO92 (homologia interna do gene *FLO8* e externa à integração de pAR2), gerando um fragmento de aproximadamente 2460 pb (Figura 16).



**Figura 16 – Confirmação da integração de pAR2 no genoma de S288c/ARY153 (*flo8Δ::HphMX4*) por PCR.** A integração de pAR2, contendo sequências parciais do alelo funcional *FLO8* de JAY291 e marca de seleção, de resistência ao antibiótico geneticina (*KanMX4*) no genoma de ARY153 repara o defeito *flo8-1* de S288c e a linhagem passa a ter um alelo funcional *FLO8*. A amplificação realizada com *primers* ARO82 (homologia interna ao gene *KanMX4* de pAR2) e ARO92 (homologia interna do gene *FLO8* e externa à integração de pAR2) gera um fragmento de aproximadamente 2460 pb. Além disso, a linhagem S888c/ARY156 (*FLO8::KanMX4*) agora é floculante. Controle negativo de amplificação com DNA de ARY153 (CN). Marcador de peso molecular “Hyperladder I” (1 kb) Biotin (M).

A inserção de pAR2 no genoma de ARY153 repara o gene *FLO8* da linhagem de laboratório e a linhagem obtida ARY156 passa a ser floculante (*Flo8p* funcional) (Figura 1b - Introdução Geral). O reparo do gene *FLO8* em S288c leva a uma linhagem floculante (Liu *et al.* 1996).

### Inserção de um marcador de seleção ligado ao gene *FLO1* de S288c

O próximo passo foi inserir um marcador de seleção, chamado de cassete *HphMX4-FLO1*, em uma região distal ao gene *FLO1* da linhagem S288c, que agora é floculante (linhagem ARY156). A realização dessa etapa foi importante para futura seleção de linhagens F2 segregantes que herdaram o *FLO1* funcional da linhagem de laboratório, após novo cruzamento da mesma com a linhagem industrial JAY291, que não possui esse alelo. O cassete *HphMX4-FLO1* foi desenvolvido e amplificado de pAG32 como um fragmento de 1,8 kb, conforme descrito em “materiais e métodos”. A inserção do cassete *HphMX4-FLO1* ocorreu no genoma da linhagem ARY156.

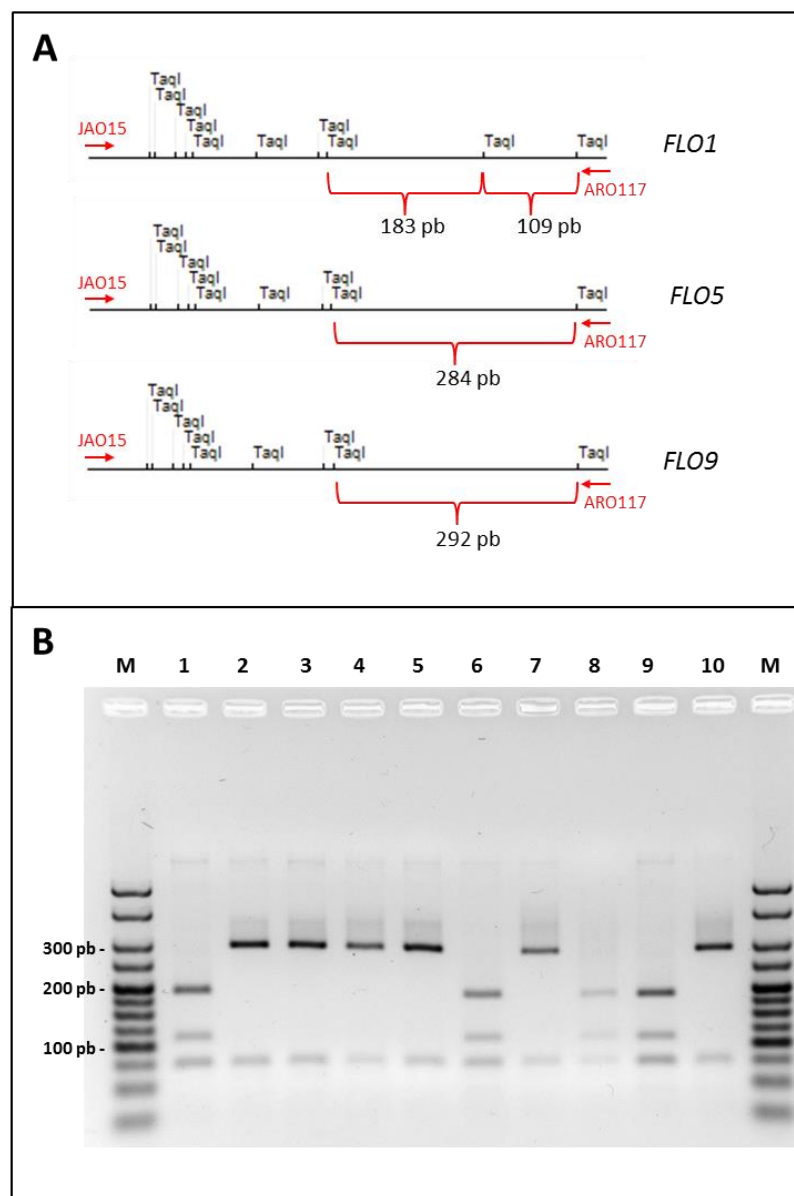


Após seleção dos transformantes em meio contendo o antibiótico higromicina e em meio contendo o antibiótico geneticina (candidatos resistentes a ambos), realizamos uma reação de PCR para confirmar a integração do cassete *HphMX4-FLO1*. Os *primers* ARO117, externo ao local alvo e à integração do cassete e JAO15, interno ao plasmídeo pAG32 e ao cassete amplificado, foram utilizados para essa confirmação (Figura 9).

Como os genes *FLO1*, *FLO5* e *FLO9* possuem altas similaridades entre si, incluindo suas regiões distais, selecionamos um sítio de restrição exclusivo da região alvo para realizarmos o PCR e analisarmos a inserção correta do cassete *HphMX4-FLO1* distal ao gene *FLO1* e não aos outros dois genes. Assim, primeiramente realizamos um PCR (*primers* ARO117-JAO15) com DNA dos transformantes candidatos e após amplificação, os fragmentos gerados de aproximadamente 600 pb foram submetidos à uma reação de digestão com a enzima de restrição *TaqI*. Se a inserção estivesse correta, ou seja, na região de *FLO1* e não de *FLO5* ou *FLO9*, o transformante deveria apresentar, após digestão com *TaqI*, duas bandas diagnósticas, uma de 109 pb e outra de 183 pb, além de outras bandas comuns às três regiões (dez sítios de restrição esperados para região distal ao gene *FLO1*). Se a inserção não estivesse correta, o fragmento teria apenas nove sítios de restrição para a enzima *TaqI* e apresentaria uma banda maior diagnóstica, não digerida, de aproximadamente 300 pb (284 pb para *FLO5* ou 292pb para *FLO9*), além das outras menores em comum com a região distal ao gene *FLO1* (Figura 17a).

Dos dez candidatos avaliados por PCR-RFLP, quatro apresentaram o padrão de restrição esperado para inserção do cassete no local desejado, distal ao gene *FLO1*, mostrando as bandas de aproximadamente 109 e 183 pb. Os outros seis candidatos

apresentaram a banda de tamanho aproximado de 300 pb, indicando que a inserção não ocorreu no local esperado (Figura 17b).



**Figura 17 - Confirmação da integração do cassete *HphMX4-FLO1* por PCR-RFLP.** Os genes *FLO1*, *FLO5* e *FLO9* possuem altas similaridades entre suas sequências. Para confirmação da inserção de marca de resistência ao antibiótico higromicina 200 pb distal a *FLO1*, selecionamos um sítio de restrição para a enzima *TaqI* (seta), específico de *FLO1*, localizado na região alvo. Realizamos o PCR com os primers JAO15 (interno ao cassete) e ARO117 (externo à integração do cassete), seguido de reação de digestão com a enzima *TaqI*. O fragmento de interesse possui dez sítios de restrição para a enzima, enquanto que se o cassete fosse inserido nas regiões distais aos genes *FLO5* e *FLO9*, os produtos das amplificações teriam apenas nove sítios (A). Em B, gel de agarose 2,5% com amostras de reação de PCR descrita em (A) realizada para dez candidatos, após restrição com enzima *TaqI*. Dos dez candidatos testados, quatro (1, 6, 8 e 9) apresentaram as bandas esperadas de aproximadamente 109 e 183 pb. Os outros candidatos não tiveram a inserção no local correto, mostrando uma banda de ~300 pb. Marcador de peso molecular “HyperLadder V” (25 pb – 500 pb) Bioline (M).

Após a pré-seleção de dois candidatos pelo padrão de restrição, os produtos gerados pelas amplificações realizadas em duas reações diferentes, uma com os *primers* ARO117 (homologia externa distal à inserção) - JAO15 (homologia interna ao cassete) e outra com os *primers* ARO116 (homologia externa proximal à inserção) - JAO14 (homologia interna ao cassete) foram sequenciados. O sequenciamento dessas regiões confirmou a inserção do cassete no local esperado, a aproximadamente 200 pb distal ao gene *FLO1*. A linhagem S288c floculante, com *FLO8* reparado e marca de seleção *HphMX4* ligada ao gene *FLO1* foi designada como ARY157.

### **Nova população F2 segregante para análise**

Para encontrarmos o terceiro gene associado ao fenótipo de floculação, a nova linhagem desenvolvida e derivada da linhagem de laboratório S288c, ARY157 (*FLO8::KanMX4 FLO1::HphMX4*), foi cruzada com a linhagem industrial JAY291 (*FLO8 flo1Δ*). O híbrido resultante não floculante foi chamado ARY161. A esporulação e consequente dissecação de tétrades de ARY161 resultaram em uma nova população F2 segregante (Figura 10).

Analisamos os fenótipos de 868 indivíduos segregantes haploides F2 de ARY161. Dos 868 indivíduos analisados, 434 possuíam resistência a higromicina, portanto possuíam *FLO1* herdado da linhagem S288c (ARY157), além do *FLO8* funcional, seja herdado de JAY291 ou de ARY157. Desses, 200 indivíduos foram classificados como floculantes e 198 como não floculantes (Tabela 3), uma relação fenotípica de 1:1, esperada para uma característica controlada por um único gene. Esse gene poderia ser o terceiro fator que esperávamos encontrar no mapeamento preliminar realizado (os três genes não ligados

esperados pela proporção fenotípica de 1:8 encontrada na população F2 segregante do mapeamento preliminar, sendo que já tínhamos encontrado *FLO1* e *FLO8* como relacionados ao fenótipo).

Tabela 3 - Fenótipo observado na população F2 segregante do cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*).

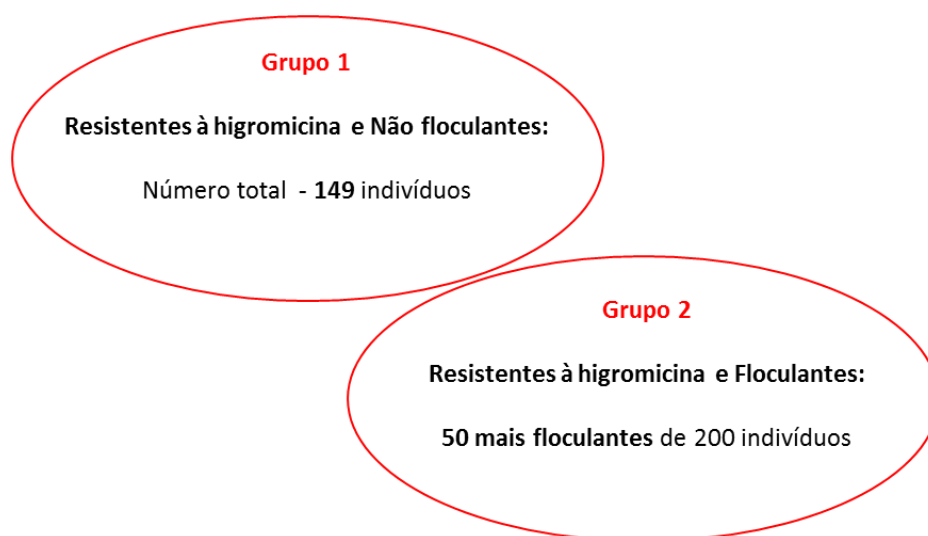
Fenótipo observado	Número de indivíduos	
	Resistentes à Higromicina	Sensíveis à Higromicina
Floculantes	200	35
Não floculantes	198	384
Indeterminado	36	15
Total de indivíduos	434	434
	868	

Porém, dentre os indivíduos floculantes, novamente pudemos perceber uma significativa diferença quanto ao grau de floculação, sendo que algumas linhagens apresentavam sedimentação imediata após a retirada do tubo do vórtex e outras levavam um tempo relativo maior para sedimentação.

Nossa hipótese era de que *FLO1*, *FLO8* e o terceiro gene esperado poderiam atuar sobre o componente discreto de herança da característica, controlando se a linhagem era floculante ou não. Por outro lado, existia também o componente quantitativo, contendo genes com efeito menor sobre a característica, que poderiam ser também atuantes sobre a variação no grau de floculação entre os segregantes.

Para o novo sequenciamento e mapeamento em “bulks”, selecionamos 149 indivíduos não floculantes e resistentes ao antibiótico higromicina como o primeiro “pool” de indivíduos (Grupo 1). Para o segundo “pool” (Grupo 2), selecionamos linhagens que apresentavam o grau máximo de floculação, comparativamente entre os 200 indivíduos

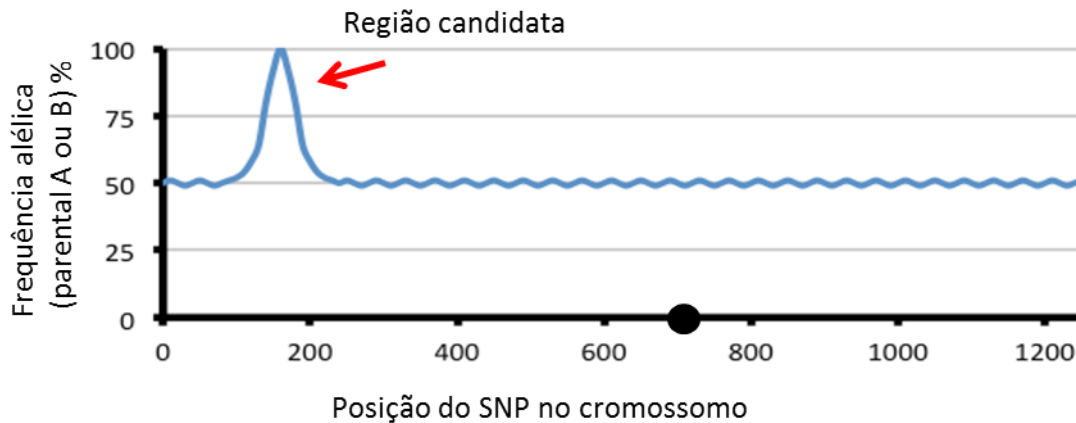
classificados como floculantes e resistentes à higromicina, resultando em um total de 50 haploides mais floculantes (Figura 18).



**Figura 18 – Ilustração dos dois “bulks” sequenciados.** Dentre a população F2 segregante entre o cruzamento de S288c e JAY291, selecionamos os indivíduos portadores do marcador de resistência à higromicina (*FLO1-HphMX4*) e dividimos em dois “bulks” para sequenciamento: 149 indivíduos não floculantes (Grupo 1) e 50 indivíduos mais floculantes (Grupo 2).

O sequenciamento foi realizado em equipamentos de última geração, conforme descrito em “materiais e métodos”, gerando coberturas médias, baseadas no mapeamento dos *reads*, de 486x para o primeiro grupo (não floculantes; resistentes à higromicina) e 891x para o segundo grupo (floculantes; resistentes à higromicina).

Para avaliarmos os resultados de sequenciamento, os *reads* gerados foram mapeados no genoma referência da linhagem S288c (disponível em SGD) e analisamos as frequências alélicas parentais S288c ou JAY291 nas sequências de cada “bulk”. Loci que afetam a característica em estudo mostram um enriquecimento para um dos parentais em cada grupo, enquanto que loci não ligados possuem segregação uniforme (Edwards and Gifford 2012), como ilustrado na Figura 19.

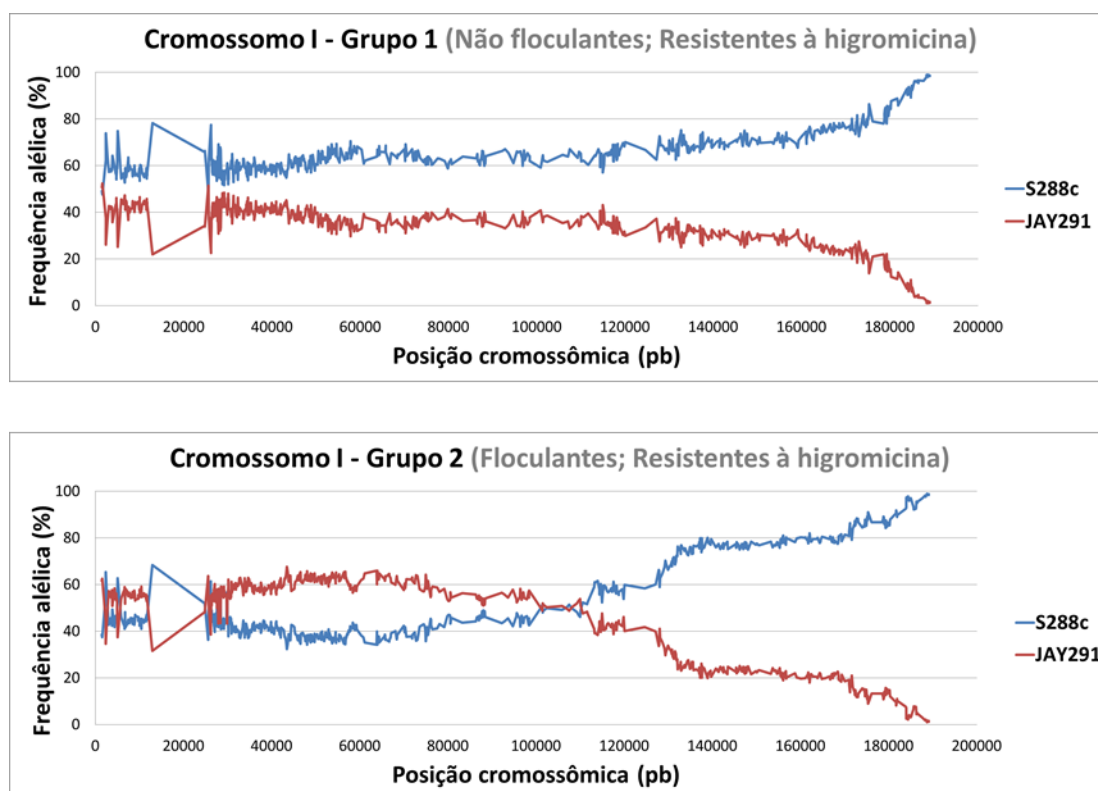


**Figura 19 - Gráfico hipotético representando o mapeamento.** A frequência alélica (%) herdada de um parental “A” ou “B” (eixo “y”) é avaliada ao longo de cada cromossomo (“eixo x”) onde estão localizadas as posições dos SNPs. Loci que afetam a característica em estudo mostram um enriquecimento de alelos para um dos parentais (seta).

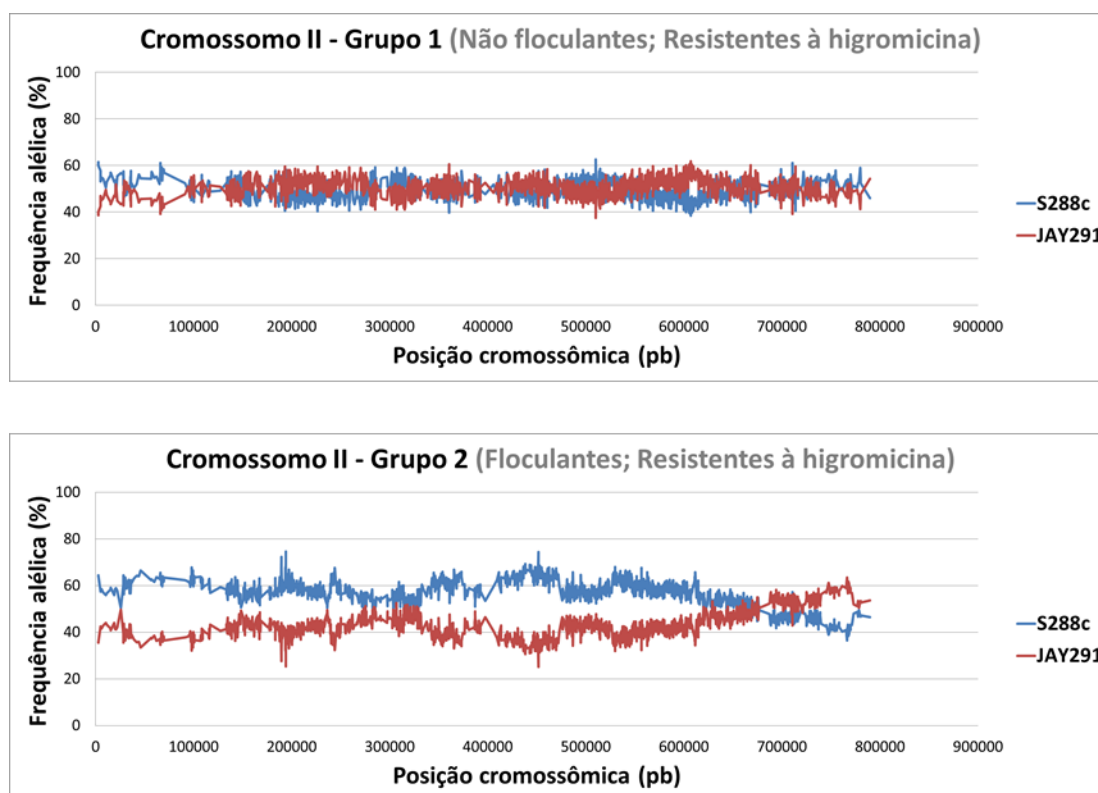
### **Mapeamento genômico: Frequência alélica parental X Posição cromossômica (SNP)**

As próximas figuras representam os resultados do mapeamento genômico, gerados pela análise gráfica das frequências alélicas parentais ao longo dos dezesseis cromossomos. Pelo número de SNPs avaliados, temos aproximadamente um SNP a cada 300 pb. Cada figura corresponde a análise de cada cromossomo individualmente, mostrando os dados gerados para o Grupo 1 (não floculantes e resistentes à higromicina) e para o Grupo 2 (50 mais floculantes e resistentes à higromicina).

Analisando graficamente os resultados, temos no “eixo x” as posições dos SNPs ao longo dos cromossomos e no “eixo y” a frequência alélica (%) parental S288c-ARY157 (linha azul) ou JAY291 (linha vermelha) (Figuras 20-35).

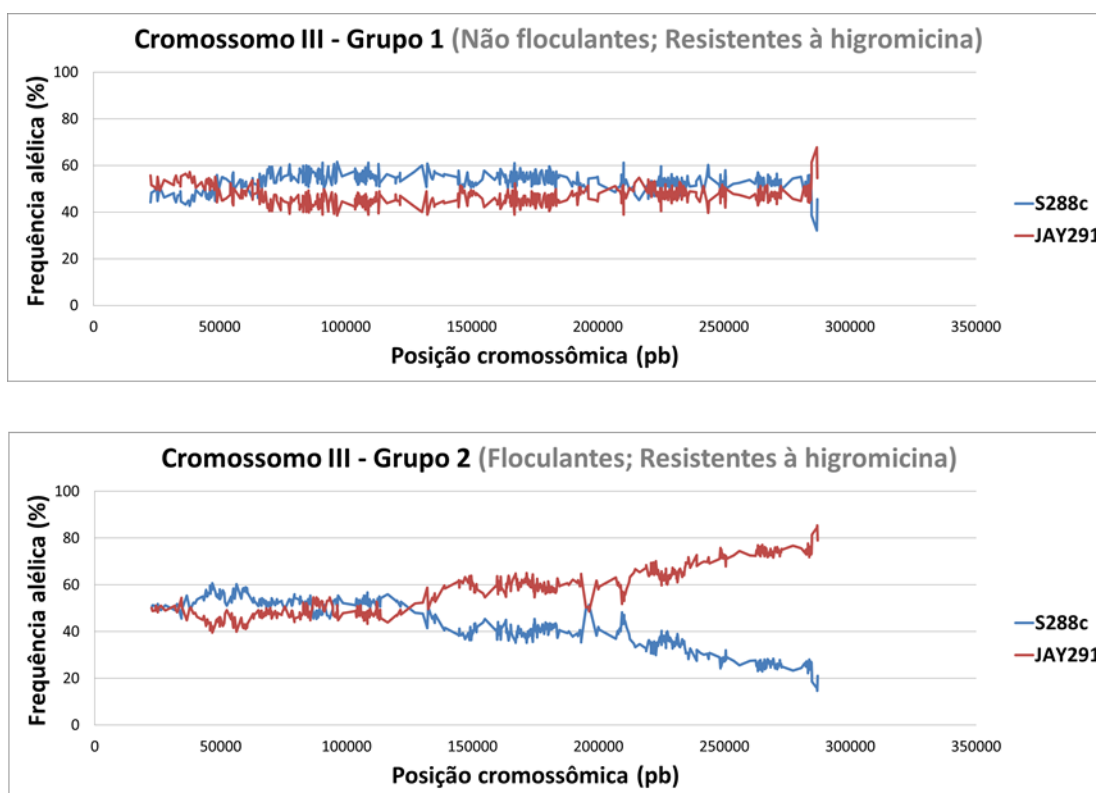


**Figura 20 - Mapeamento genético de loci envolvidos com flocculação através de sequenciamento genômico – Cromossomo I.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não flocculantes (Grupo 1) e flocculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.

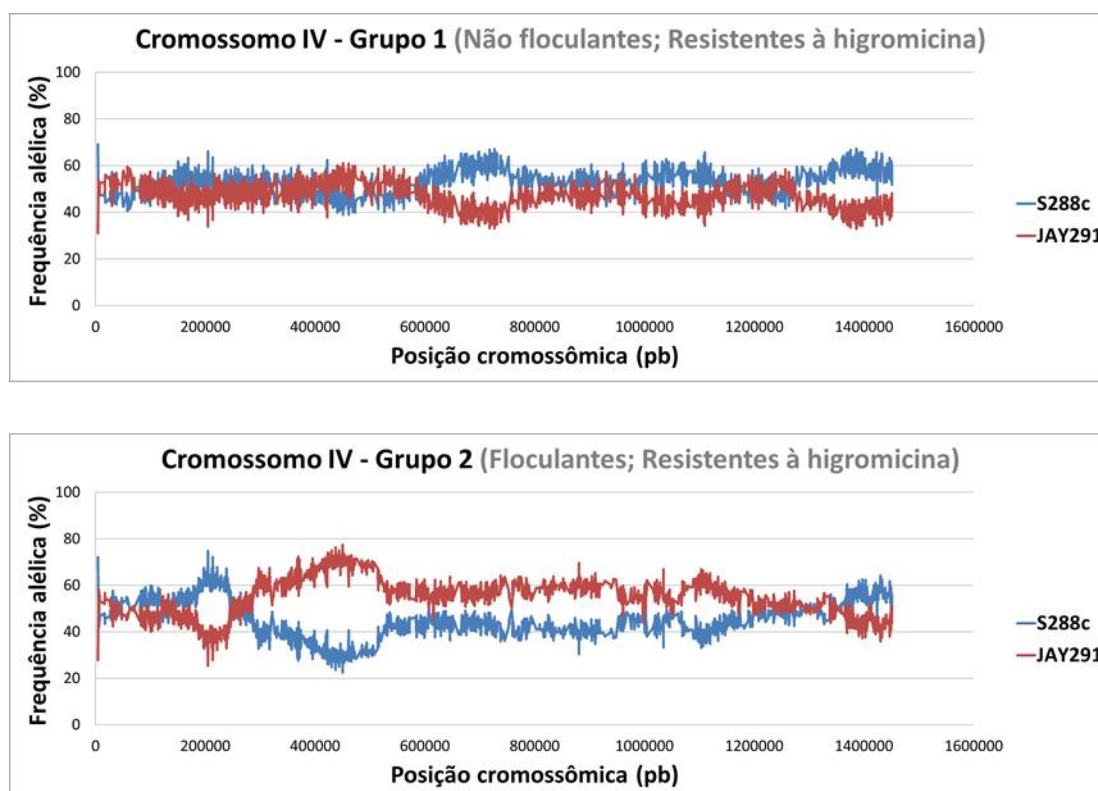


**Figura 21 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo II.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.

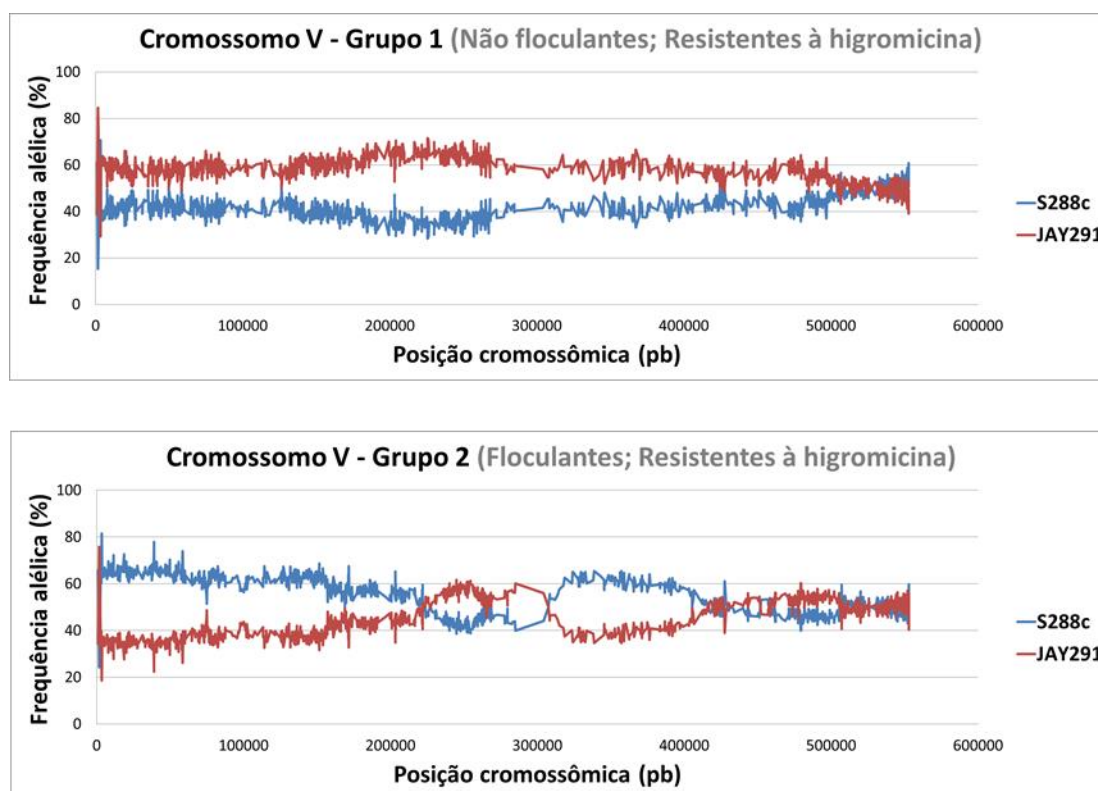




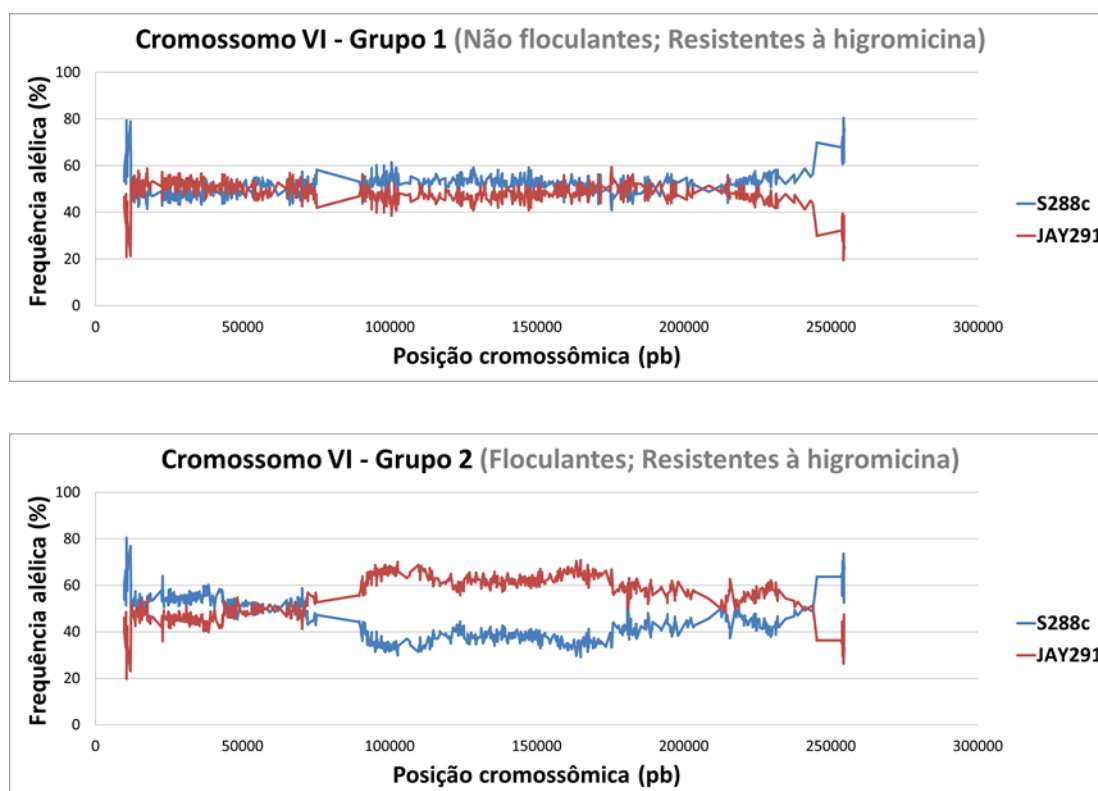
**Figura 22 - Mapeamento genético de loci envolvidos com flocculação através de sequenciamento genômico – Cromossomo III.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não flocculantes (Grupo 1) e flocculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



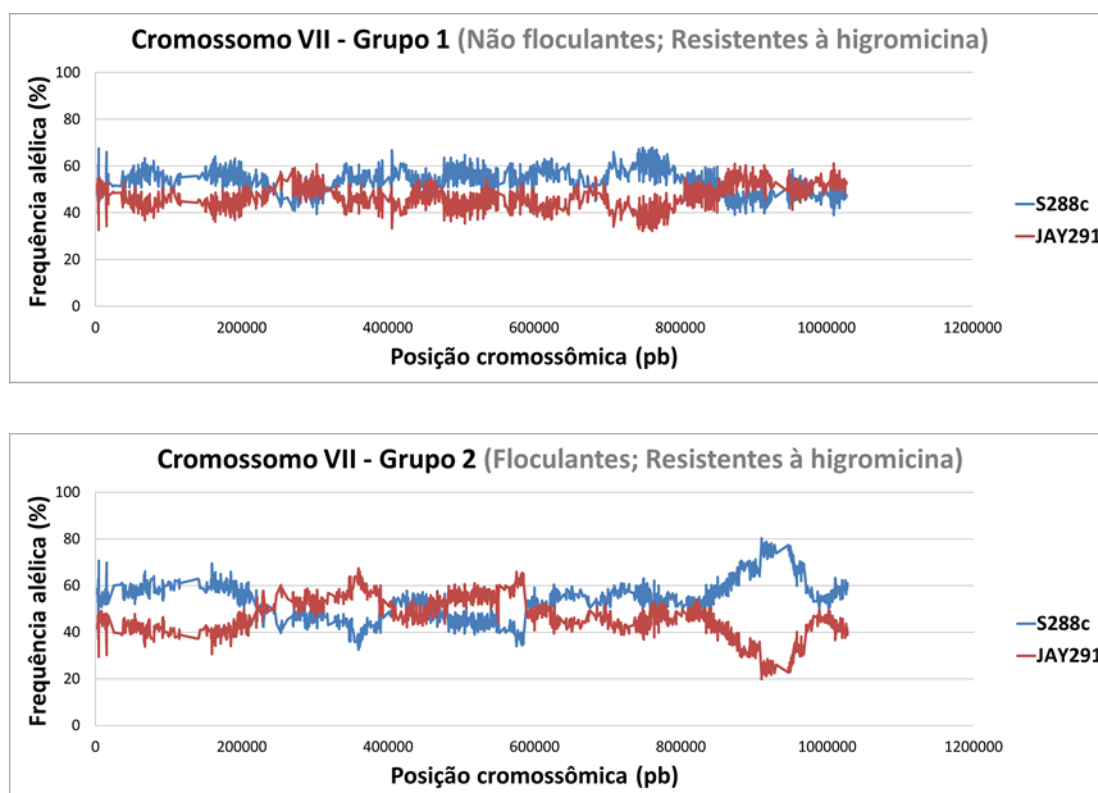
**Figura 23 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo IV.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1::HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



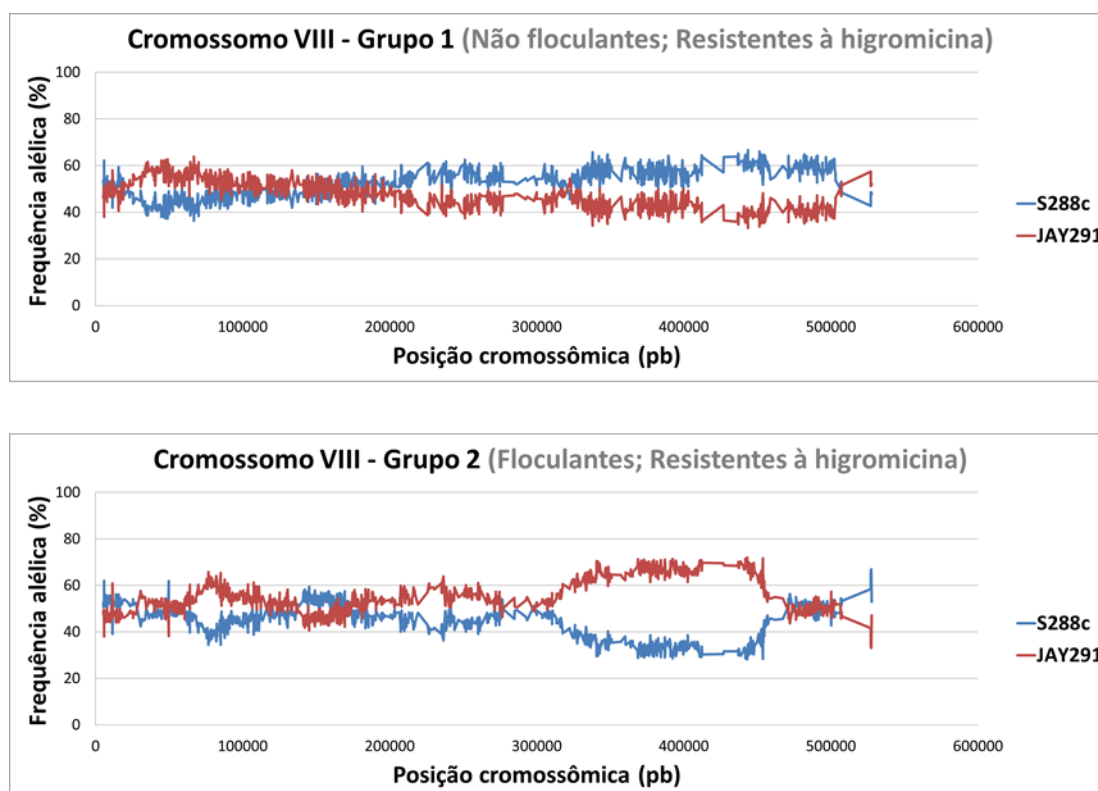
**Figura 24 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo V.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



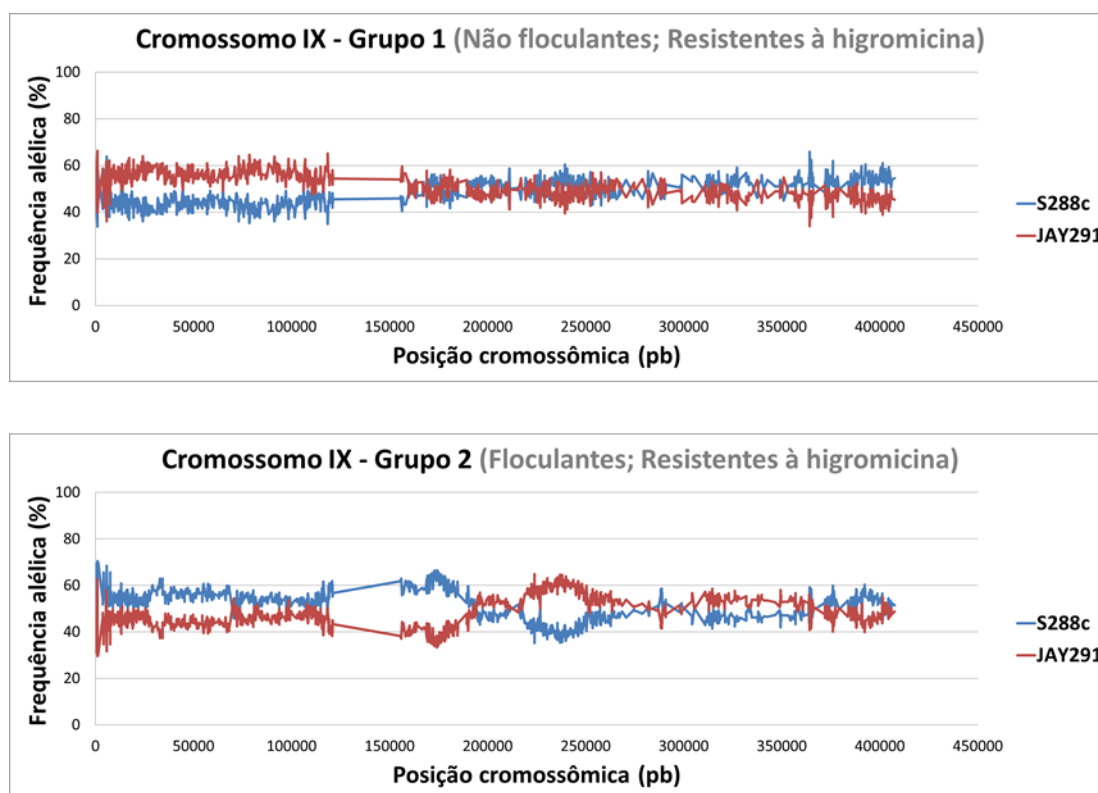
**Figura 25 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo VI.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



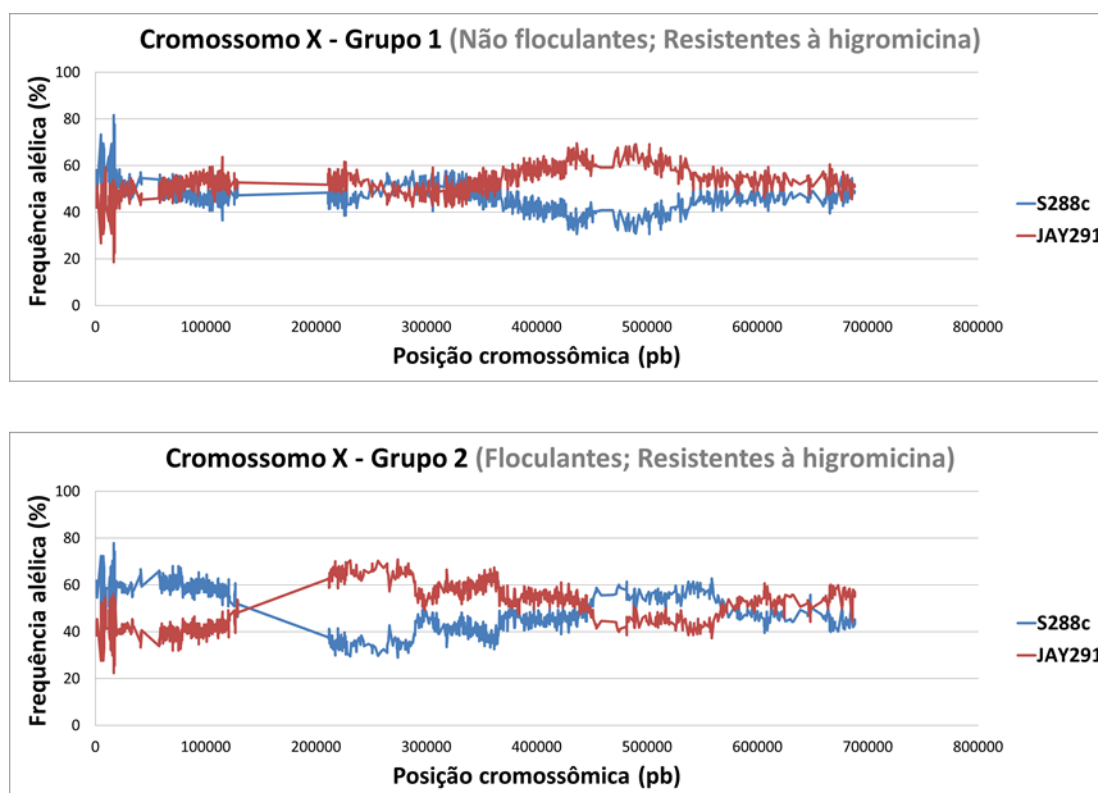
**Figura 26 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo VII.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



**Figura 27 - Mapeamento genético de loci envolvidos com flocculação através de sequenciamento genômico – Cromossomo VIII.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não flocculantes (Grupo 1) e flocculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.

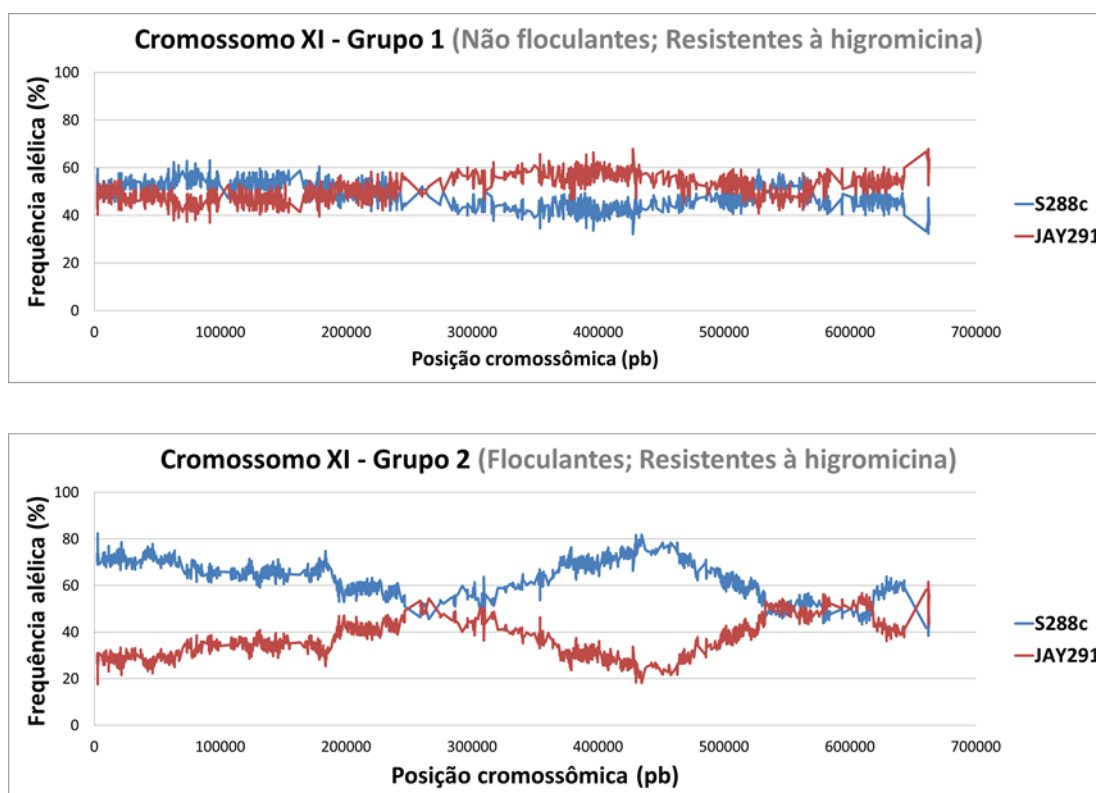


**Figura 28 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo IX.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1::HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.

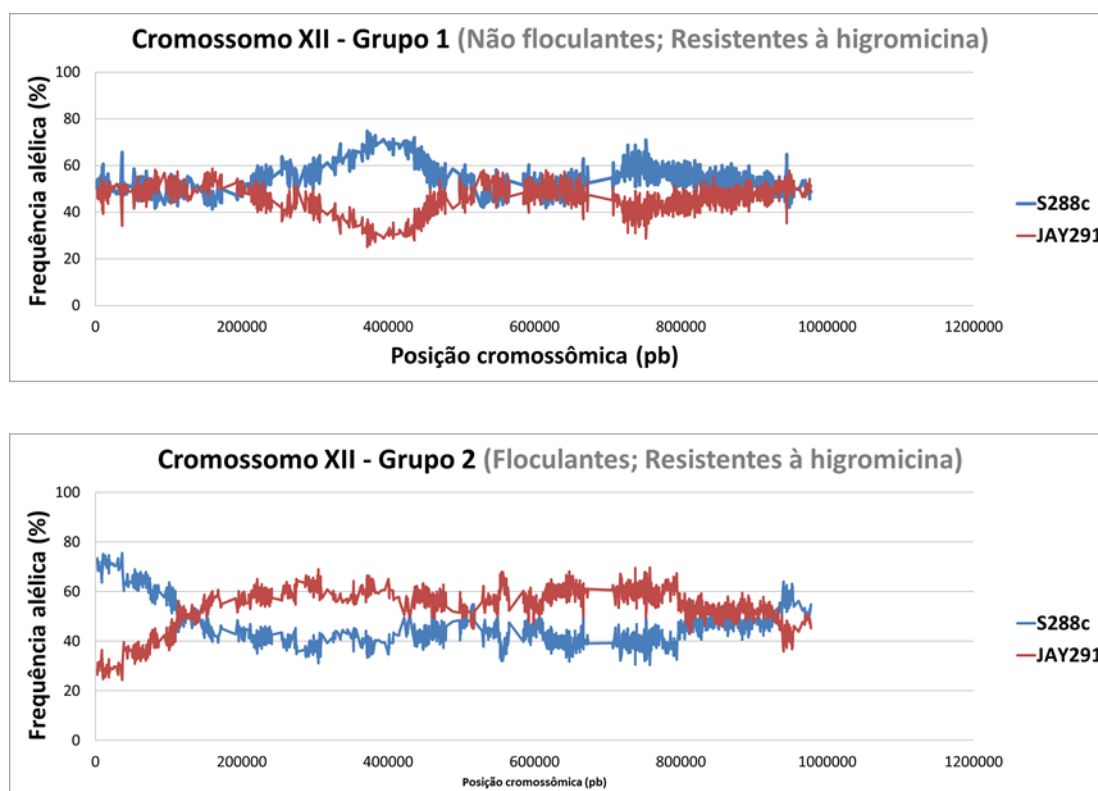


**Figura 29 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo X.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.

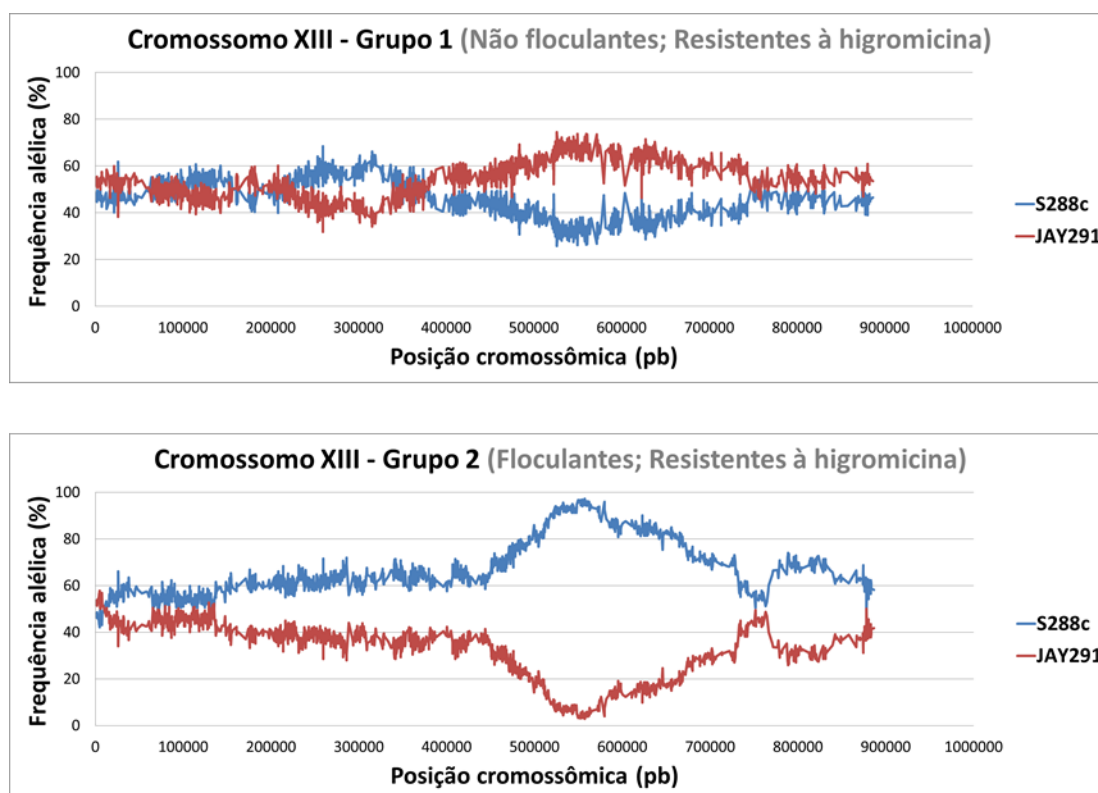




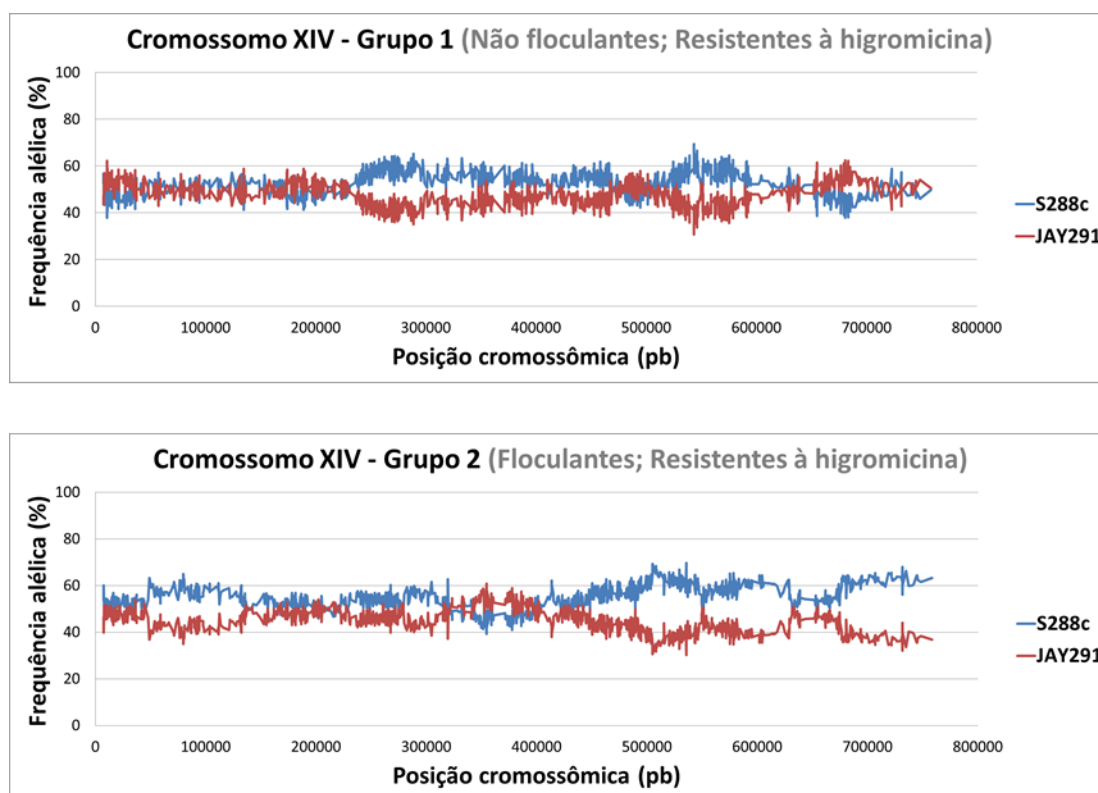
**Figura 30 - Mapeamento genético de loci envolvidos com flocculação através de sequenciamento genômico – Cromossomo XI.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não flocculantes (Grupo 1) e flocculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



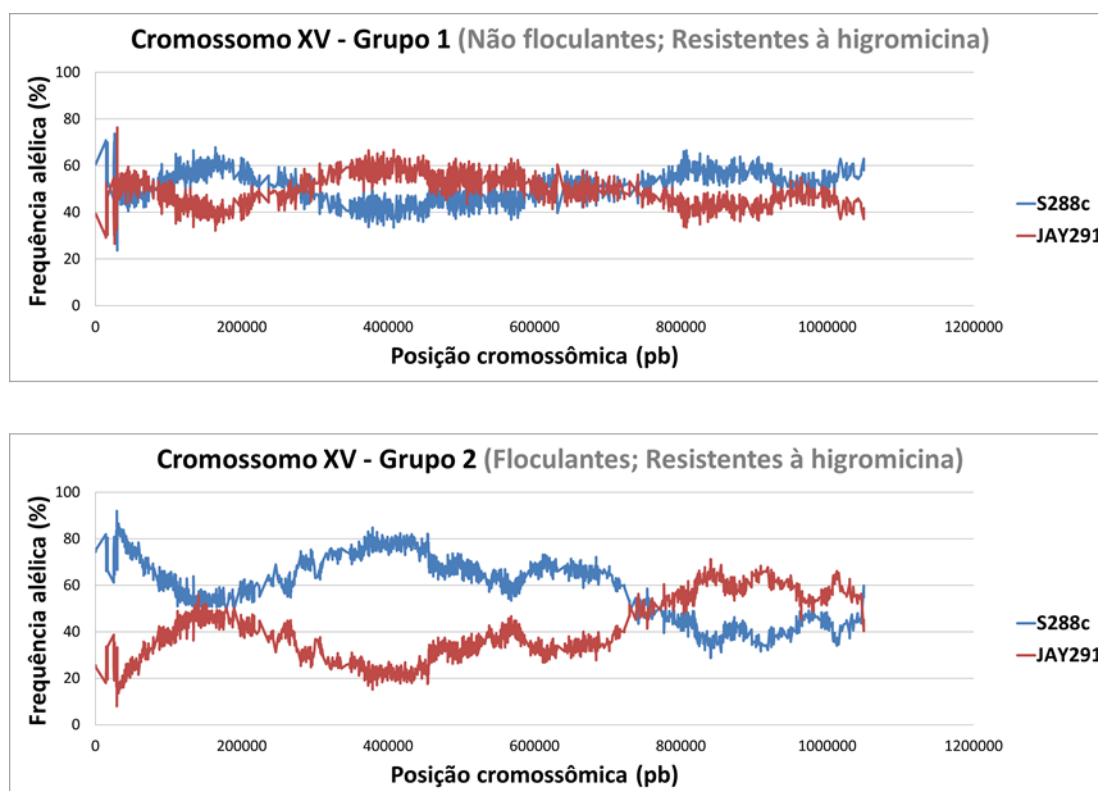
**Figura 31 - Mapeamento genético de loci envolvidos com flocculação através de sequenciamento genômico – Cromossomo XII.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não flocculantes (Grupo 1) e flocculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



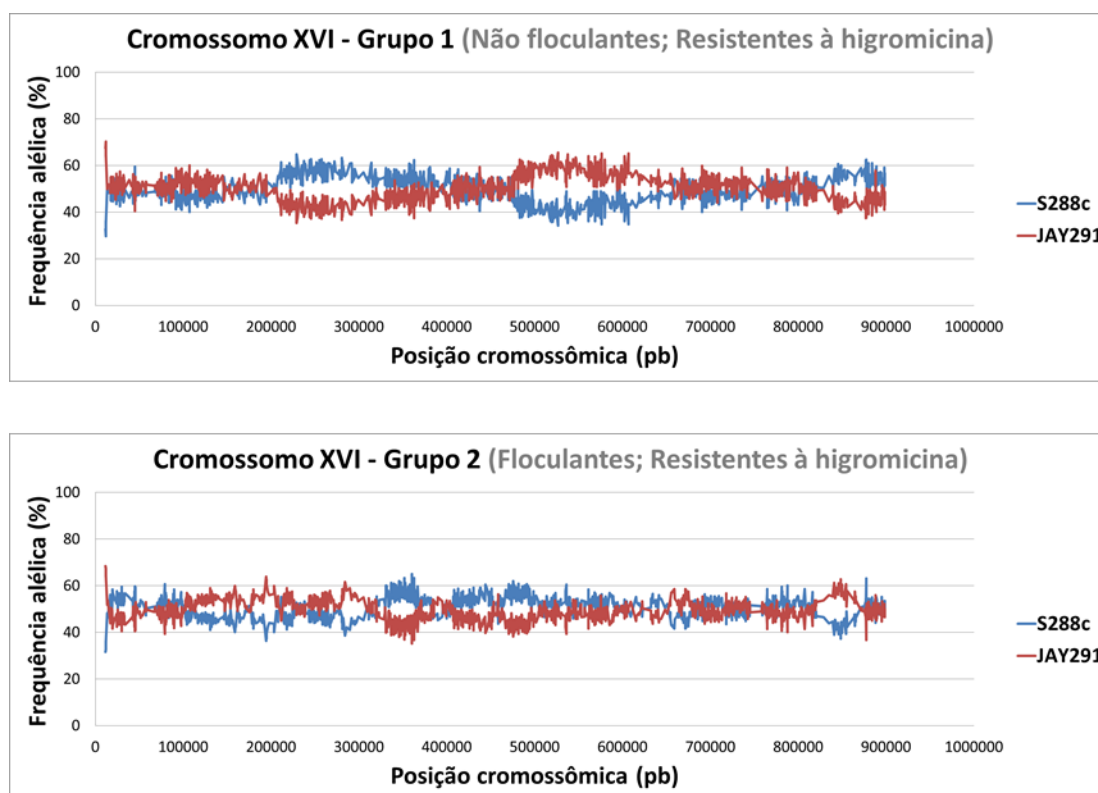
**Figura 32 - Mapeamento genético de loci envolvidos com flocculação através de sequenciamento genômico – Cromossomo XIII.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1::HphMX4* divididos em dois “bulks”: não flocculantes (Grupo 1) e flocculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



**Figura 33 - Mapeamento genético de loci envolvidos com floclação através de sequenciamento genômico – Cromossomo XIV.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1::HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



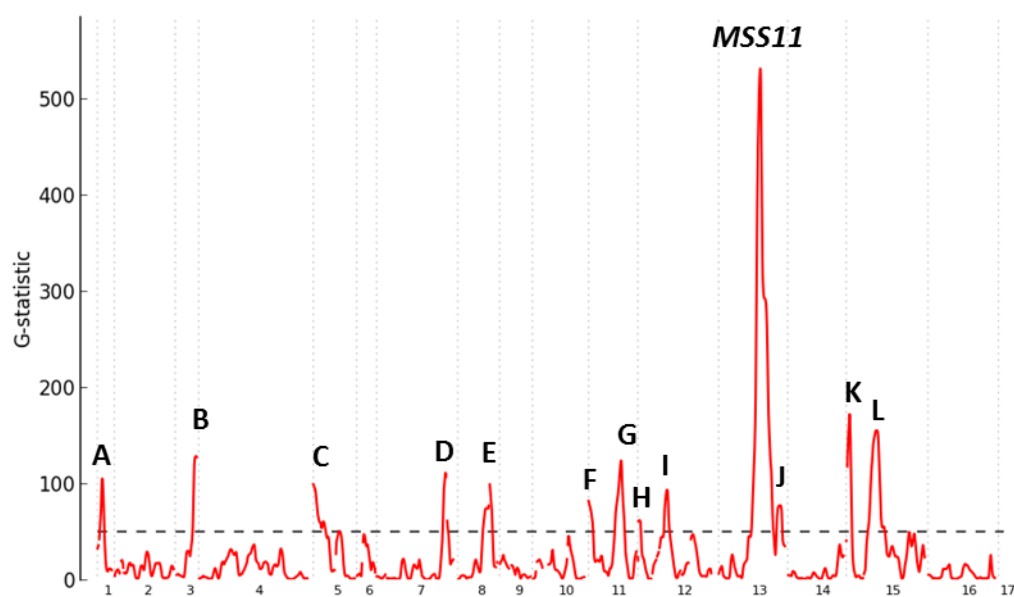
**Figura 34 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo XV.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1:HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.



**Figura 35 - Mapeamento genético de loci envolvidos com floculação através de sequenciamento genômico – Cromossomo XVI.** Haploides F2 segregantes de um cruzamento de S288c (*FLO8::KanMX4 FLO1::HphMX4*) e JAY291 (*FLO8 flo1Δ*) foram analisados fenotipicamente e indivíduos *FLO1::HphMX4* divididos em dois “bulks”: não floculantes (Grupo 1) e floculantes (Grupo 2) e sequenciados. As frequências alélicas dos parentais S288c (linha azul) e JAY291 (linha vermelha) foram interrogadas em aproximadamente 40.000 SNPs existentes entre as duas linhagens, em sequências de cada “bulk” (pool de indivíduos), que foram plotadas sobre as posições cromossômicas. Desvios significativos em relação à média de 50% indicam loci candidatos, posteriormente validados estatisticamente.

## Loci associados com o fenótipo floculante encontrados pelo mapeamento e validados estatisticamente

Através dos estudos estatísticos realizados pela comparação dos dados obtidos pelo sequenciamento (frequências alélicas em cada posição cromossômica analisada – Figuras 20-35) do Grupo 1 versus o Grupo 2 encontramos treze loci geneticamente ligados ao fenótipo floculante (Figura 36).



**Figura 36 – Treze loci significativamente associados com floculação.** A análise estatística (*G-statistic*) foi estimada por comparação entre os dados gerados pelos dois “bulks” sequenciados: Grupo 1 (não floculantes e resistentes à higromicina) versus Grupo 2 (floculantes e resistentes à higromicina). Os loci foram nomeados pelas letras de “A” a “L”. Em um dos picos localizados no Cromossomo XIII está o gene *MSS11*, fortemente associado com o fenótipo. Os números de 1 a 16 indicam cada um dos cromossomos (o começo e fim de cada cromossomo estão demarcados por linhas pontilhadas ao fundo da imagem).

A seguir, os treze loci encontrados e nomeados de “A” a “L” e “*MSS11*” na Figura 36 são apresentados na Tabela 4. Nessa tabela, na primeira coluna temos em qual cromossomo o loci está localizado e o nome do loci; a segunda coluna corresponde às

posições dos SNPs do loci no genoma da linhagem JAY291; na terceira coluna temos a correspondência dessas regiões no genoma da linhagem S288c em genes (SGD); na quarta coluna são apresentadas as posições das regiões que possuem os valores máximos estatísticos (picos - *G-statistic*) e na última coluna os genes que estão localizados nesses picos no genoma de S288c (Tabela 4).

**Tabela 4 – Treze loci associados com floculação encontrados pelo mapeamento genômico e validados estatisticamente.**

Loci	Loci - Genoma JAY291 - Posições SNPs (pb)	Loci - Genoma S288c - SGD (Genes)	Picos - Genoma JAY291 - Posições SNPs (pb)	Genes localizados nos picos - Genoma S288c - SGD
<b>Cromossomo I:</b>				
A	59192 a 73629	PTA1 a RBG1	63783 a 68062	CDC24, CLN3 e CYC3
<b>Cromossomo III:</b>				
B	267361 a 287148	CDC39 a YCR101C	277391 a 282798	GIT1
<b>Região subtelomérica</b>				
<b>Cromossomo V:</b>				
C	1066 a 21393	YEL077C a HXT13	1066 a 5333	YEL077C a YEL073C
<b>Região subtelomérica</b>				
<b>Cromossomo VII:</b>				
D	910399 a 930038	BNS1 a PFK1	922716 a 927528	YGR237C e KEL2
<b>Cromossomo VIII:</b>				
E	412350 a 446641	SOL3 a YHR177W	427086 a 436458	Entre SOL3 e SPC97
Obs: sem marcador entre 412542 e 427086 pb				
<b>Cromossomo XI:</b>				
F	2286 a 21656	YKL222C a JEN1	2286 a	YKL222C a FRE2



			9336	
		<b>Região subtelomérica</b>		
<b>G</b>	421767 a 445373	<i>UFD4</i> a <i>OSH6</i>	434455 a 439258	<i>AUR1</i> , <i>MRP17</i> e <i>DID4</i>
<b>Cromossomo XII:</b>				
<b>H</b>	7043 a 25081	<i>MMP1</i> a <i>AQY2</i>	12400 a 14026	<i>YLL058W</i>
		<b>Região subtelomérica</b>		
<b>I</b>	373396 a 395330	<i>YPS3</i> a <i>PDC5</i>	382360 a 390970	<i>DIP2</i> a <i>CKI1</i>
<b>Cromossomo XIII:</b>				
<b>MSS11</b>	545237 a 565012	<i>YMR160W</i> a <i>YMR166C</i>	556865 a 558487	<i>MSS11</i>
<b>J</b>	812963 a 831907	<i>DSS1</i> a <i>PRC1</i>	825956 a 829014	<i>HER2</i> a <i>LCB1</i>
		<b>Região subtelomérica</b>		
<b>Cromossomo XV:</b>				
<b>K</b>	33992 a 52813	<i>ZPS1</i> a <i>NOP8</i>	43801 a 45527	<i>DCP1</i> e <i>SPT20</i>
		<b>Região subtelomérica</b>		
<b>L</b>	397460 a 416213	<i>HIR2</i> a <i>RSB1</i>	401315 a 413722	<i>snR9</i> a <i>RAT1</i>

---

Logo abaixo, os loci encontrados são apresentados em outra tabela, onde levantamos genes candidatos para cada região, de acordo com funções ou fenótipos relacionados já descritos na literatura (Tabela 5).

O principal gene encontrado nesse mapeamento foi o *MSS11*, que se mostrou fortemente associado com floculação. Esse gene está localizado no Cromossomo XIII e na população floculante (Grupo 2) quase 100% de alelos (~96%) foram herdados da linhagem de laboratório S288c (Figura 32).

Tabela 5 - Genes candidatos à associação com floculação (localizados nos picos dos treze loci encontrados pelo mapeamento).

Loci	Genes candidatos	Fenótipo/Função associados na literatura
<b>Cromossomo I:</b>		
A	<i>CLN3</i>	Crescimento invasivo (Ryan <i>et al.</i> 2012)
<b>Cromossomo III:</b>		
B		
<b>Cromossomo V:</b>		
C		
<b>Cromossomo VII:</b>		
D	<i>KEL2</i>	Pseudohifas (Ryan <i>et al.</i> 2012)
<b>Cromossomo VIII:</b>		
E	<i>THP2</i>	Crescimento invasivo, pseudohifas (Ryan <i>et al.</i> 2012)
	<i>MTG2</i>	Crescimento invasivo, biofilme (Ryan <i>et al.</i> 2012)
<b>Cromossomo XI:</b>		
F		
G	<i>DID4</i>	Biofilme (Ryan <i>et al.</i> 2012)
<b>Cromossomo XII:</b>		
H	<i>YLL058W</i>	Pseudohifas (Ryan <i>et al.</i> 2012)
I	<i>ACE2</i>	Separação celular (Capítulos 3 e 4 dessa tese)
<b>Cromossomo XIII:</b>		
<i>MSS11</i>	<i>MSS11</i>	Fator de transcrição de floculação (Bester <i>et al.</i> 2006)
J	<i>HER2</i>	Biofilme, crescimento invasivo (Ryan <i>et al.</i> 2012)
	<i>YMR295C</i>	Proteína de função desconhecida localizada na periferia celular e broto (SGD)
<b>Cromossomo XV:</b>		
K		
L	<i>RAT1</i>	Crescimento invasivo (Shively <i>et al.</i> 2013)

## DISCUSSÃO

Esse estudo revelou os três principais genes envolvidos com o fenótipo floculante de indivíduos segregantes do cruzamento entre a linhagem de laboratório S288c e a linhagem industrial PE-2/JAY291: *FLO1*, *FLO8* e *MSS11*.

*FLO1* é o gene descrito como maior responsável por floculação em *S. cerevisiae* (Kobayashi *et al.* 1998; Smukalla *et al.* 2008). No trabalho de Smukalla e colaboradores (2008) linhagens ancestrais selvagens floculantes de S288c foram testadas para níveis de expressão dos genes *FLO*, correlacionando com o grau de floculação. Os resultados mostraram uma ampla variedade de floculação, de extremamente forte a quase não existente e a expressão de *FLO1* foi relacionada com forte floculação.

Observamos que a linhagem JAY291 não apresenta o gene *FLO1* em seu genoma. Linhagens que expressam o gene *FLO1* apresentam um crescimento mais lento relativamente às linhagens que não expressam, mesmo quando manose é adicionada ao meio, inibindo a floculação (Smukalla *et al.* 2008). A ausência de *FLO1* no genoma de JAY291 pode ser importante para manter o seu bom desempenho de crescimento e consequentemente, para a produção de etanol.

Por sua vez, Flo8p é um regulador positivo de floculação, atuando na regulação do gene *FLO1*. A linhagem S288c tem uma mutação *nonsense flo8-1* que a torna não floculante. Já havia sido descrito que o reparo de *FLO8* em S288c leva a um fenótipo floculante (Liu *et al.* 1996), o que foi constatado também nesse trabalho. Por outro lado, a linhagem JAY291, possui o alelo funcional desse gene, mas não flocula (Argueso *et al.*

2009). A linhagem S288c/ARY157 (*FLO8::KanMX4 FLO1::HphMX4*), utilizada no segundo cruzamento com JAY291 para mapeamento, é floculante por ter o alelo funcional de *FLO8*.

O terceiro principal gene encontrado no mapeamento, *MSS11*, está fortemente associado com floculação. Na população que possuía nível máximo de floculação observado em nosso trabalho e utilizada para mapeamento (Grupo 2), foi observado um enriquecimento para a frequência de alelos de *MSS11* da linhagem S288c em quase 100%. Estudos descritos na literatura indicam *MSS11* como um forte indutor da floculação e sugerem que *Mss11p* induz à floculação juntamente com *Flo8p*, sendo *FLO1* o gene alvo dominante dos dois fatores nesse processo e ainda, que a deleção do gene *MSS11* leva a um fenótipo não floculante (Bester *et al.* 2006). Analisando as sequências desse gene nas duas linhagens, S288c e JAY291, ambas parecem ter alelos funcionais, então ainda não sabemos quais as diferenças nucleotídicas que são importantes para o fenótipo.

Um estudo mais aprofundado de caracterização de *MSS11* sobre o fenótipo dos indivíduos estudados será interessante para chegarmos também no nível nucleotídico relevante funcionalmente. Iremos também testar se a reposição alélica de *MSS11* no genoma da linhagem floculante S288c (*FLO8::KanMX4*), através da inserção do alelo de JAY291, impede a floculação, o que poderia também justificar o fenótipo não floculante encontrado nos híbridos JAY361 e ARY161, se os três genes principais encontrados forem os únicos atuantes do componente discreto da característica, controlando se a linhagem flocula ou não.

Além dos três principais genes, encontramos também outros doze loci associados com floculação. Esses loci podem conter genes de efeito menor, mas que atuam como

modificadores do fenótipo, podendo determinar também o grau de floculação (componente quantitativo da herança).

Entre os doze loci encontrados, seis possuem localizações subteloméricas. Em estudos anteriores de nosso grupo, foi observado que a linhagem PE-2/JAY270, altamente heterozigota, possui vários polimorfismos estruturais entre cromossomos homólogos e que esses rearranjos estão confinados à periferia dos cromossomos. Também foram identificados genes presentes no genoma da linhagem PE-2/JAY291, localizados principalmente em regiões subteloméricas, que estão ausentes em S288c (Argueso *et al.* 2009).

Os principais genes envolvidos no processo de floculação são teloméricos. É importante notar que o posicionamento subtelomérico de muitos genes de adesinas não é uma coincidência. A posição relativamente isolada no genoma permite ação epigenética e frequentes eventos de recombinação sem o risco de afetar outros genes essenciais (Halme *et al.* 2004; Verstrepen and Klis 2006; Argueso *et al.* 2009).

Como os genes presentes nos seis loci subteloméricos encontrados pelo mapeamento foram analisados através do genoma de S288c (SGD), analisando essas regiões no genoma de JAY291 não descartamos a hipótese de que poderemos ainda encontrar outros genes, diferentes daqueles localizados em S288c. Porém, nessas regiões a qualidade de montagem do genoma de JAY291 é baixa, por ser uma montagem *de novo*, não mapeada em um genoma referência. As regiões subteloméricas associadas com floculação encontradas nesse trabalho serão ainda analisadas no genoma de JAY291.

Na maioria das regiões associadas, observamos ainda a presença de genes que atuam também sobre outros fenótipos de adesão, como biofilme, crescimento invasivo e filamentoso. Já foi observado que *FLO1* e *FLO11* possuem alguns mecanismos similares de regulação, mas atuam sobre fenótipos de adesão diferentes, sendo *FLO11* envolvido com crescimento filamentoso, invasivo, pseudohifas e biofilme (Palecek *et al.* 2000; Reynolds and Fink 2001; Ryan *et al.* 2012; Granek *et al.* 2013). Por exemplo, os genes *FLO8* e *MSS11* também atuam na regulação de *FLO11* (Ryan *et al.* 2012). Genes atuantes sobre outros fenótipos de adesão foram sugeridos como candidatos, porém um estudo mais aprofundado, caracterizando-os, será também necessário.

Interessantemente, em um dos loci localizados no Cromossomo XII (loci I) foi identificado um gene que participa no processo de separação celular, *ACE2*. Observamos que nessa região as linhagens não floculantes (Grupo 1) possuem aproximadamente 70% de alelos herdados de S288c e 30% de JAY291, enquanto que as floculantes (Grupo 2) possuem aproximadamente 40% de alelos herdados de S288c e 60% de alelos de JAY291. Em estudos não relacionados com esse capítulo, que serão abordados nos capítulos 3 e 4 dessa tese, observamos que JAY270 é heterozigota para uma mutação *frameshift* em *ACE2* (*ACE2/ace-7A*), sendo que a linhagem JAY291 é portadora do alelo mutado *ace-7A*. Linhagens derivadas de JAY270 homozigotas para *ace-7A* (*ace-7A/ace-7A*) apresentaram alteração no fenótipo de colônia e baixo nível de floculação.

## CONCLUSÕES

Através de mapeamento genômico em indivíduos F2 floculantes segregantes do cruzamento da linhagem de laboratório S288c e industrial PE-2/JAY291, concluímos que:

- ✓ Os genes *FLO1*, *FLO8* e *MSS11* são necessários para que a floculação ocorra;
- ✓ Existem mais doze loci associados com floculação que contém genes de efeito menor que atuam como modificadores do fenótipo;
- ✓ Muitos dos loci encontrados estão localizados em regiões subteloméricas;
- ✓ Foram dados os primeiros passos para o entendimento da genômica de PE-2/JAY291 que está associada ao seu fenótipo não floculante, estudos adicionais serão necessários para a completa compreensão da complexa rede envolvida com floculação.

## REFERÊNCIAS

- Argueso, J. L., M. F. Carazzolle, P. A. Mieczkowski, F. M. Duarte, O. V. C. Netto *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19: 2258–70.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.*, 1998 *Current Protocols in Molecular Biology*. John Wiley & Sons Inc., New York.
- Bester, M. C., I. S. Pretorius, and F. F. Bauer, 2006 The regulation of *Saccharomyces cerevisiae* *FLO* gene expression and Ca<sup>2+</sup> -dependent flocculation by Flo8p and Mss11p. *Curr. Genet.* 49: 375–83.
- Carvalho-Netto, O. V., A. Rodrigues, L. V. Santos, M. F. Carazzolle, N. M. V. Sampaio *et al.*, 2012 Biotecnologia em leveduras, pp. 371–400 in *Biotecnologia Aplicada à Agropecuária.*, edited by G. M. de A. Cançado and L. N. Londe. EPAMIG Sul de Minas, Caldas, MG.
- Connelly, C. F., and J. M. Akey, 2012 On the prospects of whole-genome association mapping in *Saccharomyces cerevisiae*. *Genetics* 191: 1345–53.

- Edwards, M. D., and D. K. Gifford, 2012 High-resolution genetic mapping with pooled sequencing. *BMC Bioinformatics* 13 Suppl 6: S8.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1996 Life with 6000 genes. *Science* 274: 546, 563–7.
- Goldstein, a L., and J. H. McCusker, 1999 Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 15: 1541–53.
- Granek, J. a, D. Murray, Ö. Kayrkçi, and P. M. Magwene, 2013 The genetic architecture of biofilm formation in a clinical isolate of *Saccharomyces cerevisiae*. *Genetics* 193: 587–600.
- Halme, A., S. Bumgarner, C. Styles, and G. R. Fink, 2004 Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. *Cell* 116: 405–15.
- Huxley, C., G. ED, and I. Dunham, 1990 Rapid assessment of *S. cerevisiae* mating type by PCR. *Trends Genet.* 6: 236.
- Kobayashi, O., N. Hayashi, R. Kuroki, and H. Sone, 1998 Region of *FLO1* proteins responsible for sugar recognition. *J. Bacteriol.* 180: 6503–10.
- Liti, G., and E. J. Louis, 2012 Advances in quantitative trait analysis in yeast. *PLoS Genet.* 8: e1002912.
- Liu, H., C. A. Styles, and G. R. Fink, 1996 *Saccharomyces cerevisiae* S288C has a mutation in *FL08*, a gene required for filamentous growth. *Genetics* 144: 967–978.
- Magwene, P. M., J. H. Willis, and J. K. Kelly, 2011 The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput. Biol.* 7: e1002255.
- Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454: 479–85.
- Palecek, S. P., a S. Parikh, and S. J. Kron, 2000 Genetic analysis reveals that *FLO11* upregulation and cell polarization independently regulate invasive growth in *Saccharomyces cerevisiae*. *Genetics* 156: 1005–23.
- Reynolds, T. B., and G. R. Fink, 2001 Bakers' yeast, a model for fungal biofilm formation. *Science* 291: 878–81.
- Rose, M. D., F. Winston, and P. Hieter, 1990 *Methods in yeast genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.



- Ryan, O., R. S. Shapiro, C. F. Kurat, D. Mayhew, A. Baryshnikova *et al.*, 2012 Global gene deletion analysis exploring yeast filamentous growth. *Science* 337: 1353–6.
- Shively, C. a, M. J. Eckwahl, C. J. Dobry, D. Mellacheruvu, A. Nesvizhskii *et al.*, 2013 Genetic networks inducing invasive growth in *Saccharomyces cerevisiae* identified through systematic genome-wide overexpression. *Genetics* 193: 1297–310.
- Smukalla, S., M. Caldara, N. Pochet, A. Beauvais, C. Yan *et al.*, 2008 *FLO1* is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* 135: 726–737.
- Steinmetz, L. M., H. Sinha, D. R. Richards, J. I. Spiegelman, P. J. Oefner *et al.*, 2002 Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416: 326–30.
- Swinnen, S., J. M. Thevelein, and E. Nevoigt, 2012 Genetic mapping of quantitative phenotypic traits in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12: 215–27.
- Verstrepen, K. J., and F. M. Klis, 2006 Flocculation, adhesion and biofilm formation in yeasts. *Mol. Microbiol.* 60: 5–15.
- Wach, A., A. Brachat, R. Pöhlmann, and P. Philippsen, 1994 New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 10: 1793–1808.



## Capítulo 2

**Colony morphology transition in industrial yeast strains is associated with mitotic recombination leading to loss-of-heterozygosity**

Esse capítulo contém dados e experimentos de um estudo em fase final de redação de manuscrito, o qual será submetido à revista *Genetics* em breve. Sou a primeira e principal autora nesse manuscrito.



# **Colony morphology transition in industrial yeast strains is associated with mitotic recombination leading to loss-of-heterozygosity**

## **INTRODUCTION**

The fermentation process for bioethanol production used by Brazilian distilleries occurs in large tanks containing high yeast cell densities. At the end of the fermentation, yeast cells are separated from the fermented must by centrifugation and recycled after subsequent treatment with sulfuric acid to reduce bacterial contamination. The recycling is repeated for 400 to 600 cycles during sugarcane harvest season and yeast cells are required to survive many different stressful conditions imposed during this long process in the industrial environment (Wheals *et al.* 1999; Amorim *et al.* 2011).

The *Saccharomyces cerevisiae* PE-2/JAY270 strain is widely used for bioethanol production in Brazil because it has several desirable industrial properties such as low foam formation, tolerance to stress conditions, dominance in competition against yeast contaminants, high ethanol yield, high rate of cell viability, and does not display flocculation (Basso *et al.* 2008).

Flocculation is a process of multicellular aggregation, which is calcium dependent, and is mediated by adhesin proteins of the cell wall called flocculins. This phenotype is controlled mainly by genes belonging to the *FLO* family, *FLO1*, *FLO5*, *FLO9* and *FLO10* (Halme *et al.* 2004; Verstrepen and Klis 2006; Van Mulders *et al.* 2009; Soares 2011; Honigberg 2011). Flocculation impairs the process of bioethanol production because it reduces the cellular surface contact with the substrate, reducing the fermentative kinetics

and overall yield. It can also lead to problems at the centrifugation step, and cause clogging of pipes. The presence of indigenous strains with rough colony morphology has been observed during the fermentation, frequently associated with pseudohyphal growth and a high sedimentation rate, causing problems analogous to those seen in flocculent strains (Basso *et al.* 2008). Rough colonies were also noticed in 0–10% of colonies of selected industrial strains, including PE-2, during their implantation in industrial fermentations (Basso *et al.* 2008), however, there are no studies investigating the specific genetic factors involved in the colony morphology transition observed in bioethanol yeast strains.

In this study, we isolated from laboratory cultures spontaneous derivatives of JAY270 displaying altered colony morphology, exhibiting roughened surfaces, irregular edges and a resembling flocculation sedimentation phenotype in liquid media. We identified *ACE2* as the gene responsible for this phenotype and found that JAY270 is heterozygous for a defective allele of this gene (*ACE2/ace2-7A*). Our findings demonstrated that this phenotype appeared in JAY270 through copy-neutral mitotic recombination leading to loss-of-heterozygosity (LOH).

## **MATERIALS AND METHODS**

### **Media, growth conditions, strains and microbiological techniques**

YPD, Luria-Bertani (LB), Sporulation and Minimal Media were prepared as described in standard protocols (Ausubel *et al.* 1998). Yeast strains were grown in YPD or minimal media at 30°C. Sigma-Aldrich Yeast Synthetic Drop-out Media Supplement without Uracil was used to prepare minimal media to select transformants that have

integrated pAR1 plasmid. 5-fluoro-orotic acid (5-FOA) was supplemented to 1 g/l final concentration (Boeke *et al.* 1984) to select strains that have lost pAR1. Yeast strains are listed in Table 1. All PE-2 strains are descendants or derived from the diploid JAY270, pure isolate of this strain and previously molecular genetic analyzed (Argueso *et al.* 2009). For tetrad dissection, strains were plated on sporulation medium and incubated for three days at 30°C. After sporulation, tetrads were dissected by micromanipulation. The haploid and diploid mating types of strains, *MATa*, *MATα* or *MATa/α* were checked by PCR, based on the technique described (Huxley *et al.* 1990).

**Table 1 - Yeast strains used in this study**

Strains:	Genotypes:	Descriptions:	References:
<b>Diploids:</b>			
JAY270:	$\frac{MAT\alpha}{MATa}$ $\frac{ACE2}{ace2}$	PE-2 wild-type isolate	Argueso <i>et al.</i> 2009
JAY663, JAY664, JAY665, JAY912, JAY913:	$\frac{MAT\alpha}{MATa}$ $\frac{ace2}{ace2}$	Wild-type strains derived from JAY270: spontaneous variation in colony morphology	This study
<b>Haploids:</b>			
JAY289, JAY290, JA291, JAY292:	Check figure 2	Tetrad from JAY270	Argueso <i>et al.</i> 2009
JAY293, JAY294, JAY295, JAY296:	Check figure 2	Tetrad from JAY270	Our collection
JAY325, JAY326:	Check figure 2	Tetrad from JAY270	Our collection

JAY327,  
JAY328:

ARY51,  
ARY52,  
ARY53,  
ARY54:

Check figure 2

Tetrad from JAY663

This study

ARY55,  
ARY56,  
ARY57,  
ARY58:

Check figure 2

Tetrad from JAY663

This study

ARY59,  
ARY60,  
ARY61,  
ARY62:

Check figure 2

Tetrad from JAY663

This study

## Analysis of the *ACE2* region

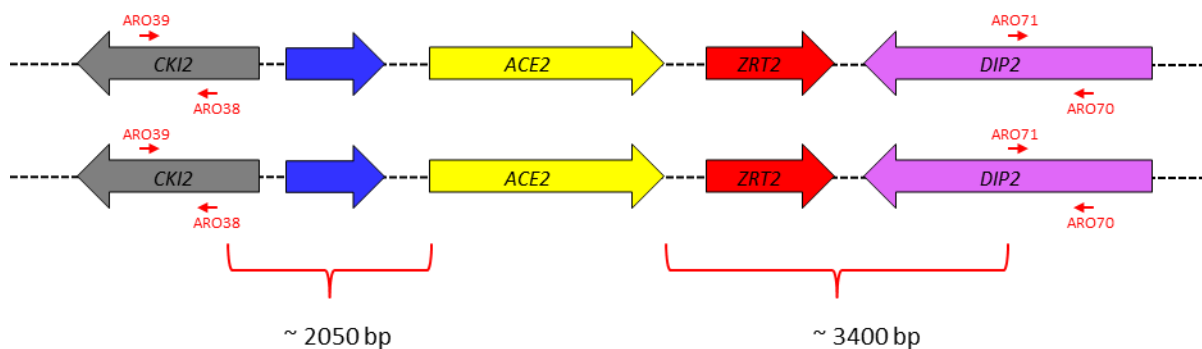
The *ACE2* gene was analyzed by PCR sequencing of three different genomic regions, using DNA samples of each diploid JAY585, JAY663, JAY664, JAY665, JAY912, JAY913 strains. All PCR amplifications were performed with “MyTaq DNA Polymerase” (Bioline). The first PCR fragment was amplified with ARO32 and ARO33 primers and contains the wild type or the mutant region identified in *ACE2* gene. The other two regions sequenced are located proximal and distal to the *ACE2* gene and each one contains Single Nucleotide Polymorphisms (SNPs) identified in the heterozygous JAY270 genome. The proximal region was amplified with ARO38 and ARO39 primers and the distal region with ARO70 and ARO71 oligonucleotides (Figure 1). We also sequenced the PCR product generated by ARO32 and ARO33 of the haploid JAY290 and JAY291 strains. Before preparation for sequencing PCR products were gel-purified with “GeneJET Gel Extraction Kit” (Thermo Scientific). Sequencing was carried out on an “ABI 3130xL Genetic Analyzer”



(Applied Biosystems) by “Proteomics and Metabolomics Facility” from Colorado State University. Sequences were uploaded and analyzed with “CLC Genomics Workbench” software. Oligonucleotide primers used in this study are listed in Table 2.

**Table 2 - Oligonucleotide primers used in this study**

Oligos:	5' → 3' sequence	Descriptions:
ARO32	TACTTCAATAAATGGTTCAC	Forward primer - <i>ACE2</i> fragment for sequencing
ARO33	TGATATTGTCGAGACCGTGG	Reverse primer - <i>ACE2</i> fragment for sequencing
ARO38	TAGCACAAGATATGAACT	Forward primer - proximal <i>ACE2</i> region
ARO39	TATATCATCTACAATGAC	Reverse primer - proximal <i>ACE2</i> region
ARO70	ATACTGGAGAGTGTTGGG	Forward primer - distal <i>ACE2</i> region
ARO71	TCTAGCGACCAAATTGCC	Reverse primer - distal <i>ACE2</i> region

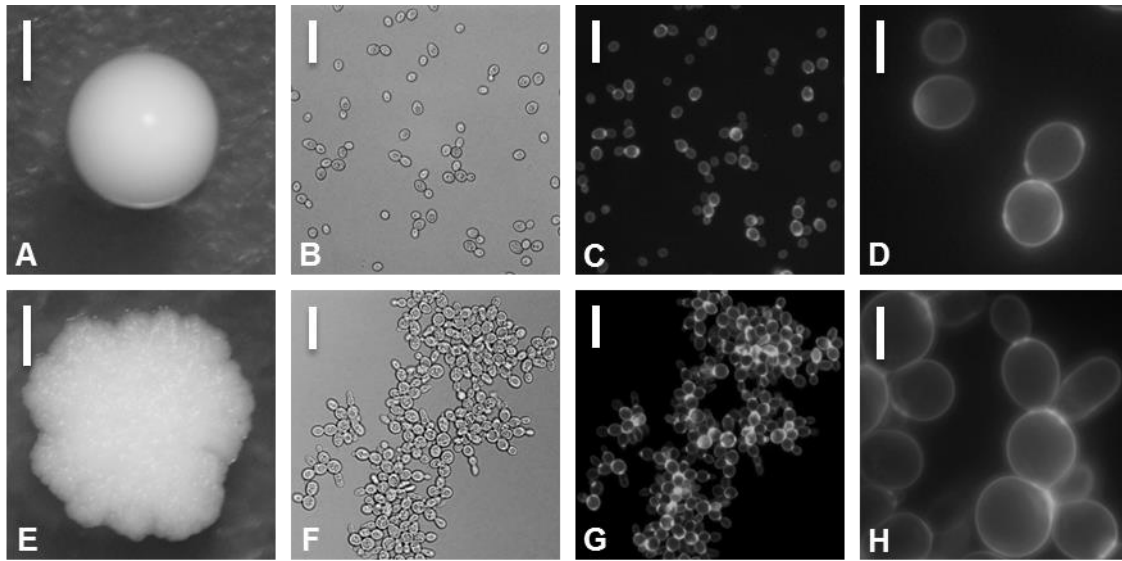


**Figure 1 - Proximal and distal *ACE2* regions analyzed in this study selected from the heterozygous JAY270 genome.** Fragment amplified with ARO38 and ARO39 oligonucleotides is located approximately at 2050 bp proximal *ACE2*. PCR product amplified with ARO70 and ARO71 primers is located at about 3400 bp distal *ACE2*

## RESULTS

### **Observation of rough colony morphology among derivatives of the JAY270 strain**

One of the most desirable features of the JAY270/PE-2 bioethanol production strain is that it does not normally flocculate during industrial sugarcane extract fermentation. Accordingly, this strain remains in suspension in liquid culture, and produces normal hemispherical colonies with smooth surfaces and edges when grown in solid agar rich medium (Figure 2A). While this is the phenotype typically observed, over the course of our studies with this strain we noticed the sporadic occurrence of colonies displaying altered morphology: relatively flat-growing colonies with rough surfaces and edges (Figure 2E). Under bright field microscopic examination, yeast cells derived from such rough colonies appeared to grow in chains, showing a budding pattern consistent with a defect or delay in the release of the daughter cells from their mother. We stained these cells with calcofluor white and used its fluorescence to highlight the chitin-rich ring septa, confirming the attachment of mother and daughter cells at the budding neck site (Figure 2).



**Figure 2 - PE-2 colony morphology variation.** Diploid wild-type JAY270 (A, B, C, D): PE-2 typical hemispherical smooth colonies (A), cells exhibiting an independent type of growing under a microscopic view (B, C, D) and stained with Calcofluor White (C, D). Diploid wild-type JAY663 (E, F, G, H): PE-2 spontaneous altered colony morphology, exhibiting irregular edges and roughened surfaces (E), chains of cells, growing attached to each other (F) and fluorescence microscopy (G, H), cells stained with CW, highlighting chitin septa and the defect in cell separation. Scale bars are 1mm (A, E), 20µm (B, C, F, G) and 5µm (D, H).

Whenever this phenotype was observed, we isolated such clones and stored them for later genetic characterization. Overtime we collected five independent examples (strains JAY663, JAY664, JAY665, JAY912 and JAY913), all of which were derived from pure cultures of JAY270-isogenic diploids. The phenotype of these strains was stably maintained over several clonal generations, suggesting that it was likely hard-wired genetically, as opposed to being caused by a transient altered transcriptional state. We estimated that the rough colonies appeared at a frequency of about one in every 10,000 colonies growing in rich media plates. Assuming a genetic origin and based on its high frequency of occurrence, we reasoned that this phenotype was unlikely to be caused by a rare *de novo* nucleotide point mutation, but instead, mitotic recombination leading to loss-of-heterozygosity (LOH) provided a more plausible explanation.

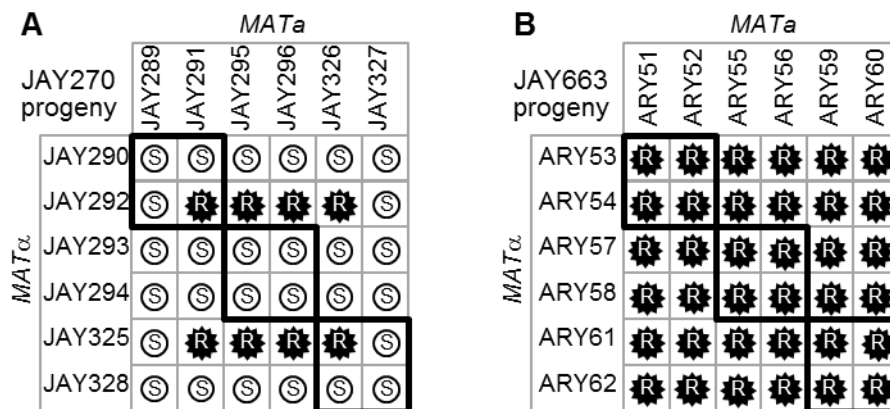
In a parallel but unrelated project, we observed that crossing two specific haploid descendants of JAY270 (JAY291 *MATa* and JAY292 *MATα*) resulted in diploid cells with the same rough colony morphology and mother-daughter cell attachment pattern observed in the spontaneous JAY270 derivatives. This result was unexpected because JAY291, JAY292, and all other haploid derivatives of JAY270 have the normal smooth colony phenotype. This fortuitous observation provided an indication that the rough colony phenotype might be diploid-specific, and the ability to consistently reproduce the mutant phenotype in a controlled cross between well-characterized haploids opened an avenue to investigate its genetic basis.

We previously reported the whole genome sequence of JAY291 (Argueso *et al.* 2009), and recently conducted the genome sequencing of JAY292 as well as of ten additional JAY270-derived haploids. This genome sequence dataset, comprising three sets of four-spore tetrads, was generated in a project to map the distribution and phasing of heterozygous loci in the JAY270 genome, the results of which will be described elsewhere (F. M. Duarte, T. M. Gurol, A. Rodrigues, *et al. in preparation*).

We carried out crosses between the 12 sequenced haploid descendants of JAY270, as shown in Figure 3A. All possible *MATa* x *MATα* crosses were performed producing in 36 different diploids. Among them, we found 8 with rough and 28 with smooth colony surfaces, a 1:4 phenotypic ratio consistent with recessive inheritance of a trait controlled by a single gene. Even though the rough colony phenotype was not observed in the haploids, the phenotypes of their respective diploid combinations allowed us to infer

which allele was present in the parents: either the wild type dominant allele or the recessive mutant allele (temporarily named *R* and *r*, respectively).

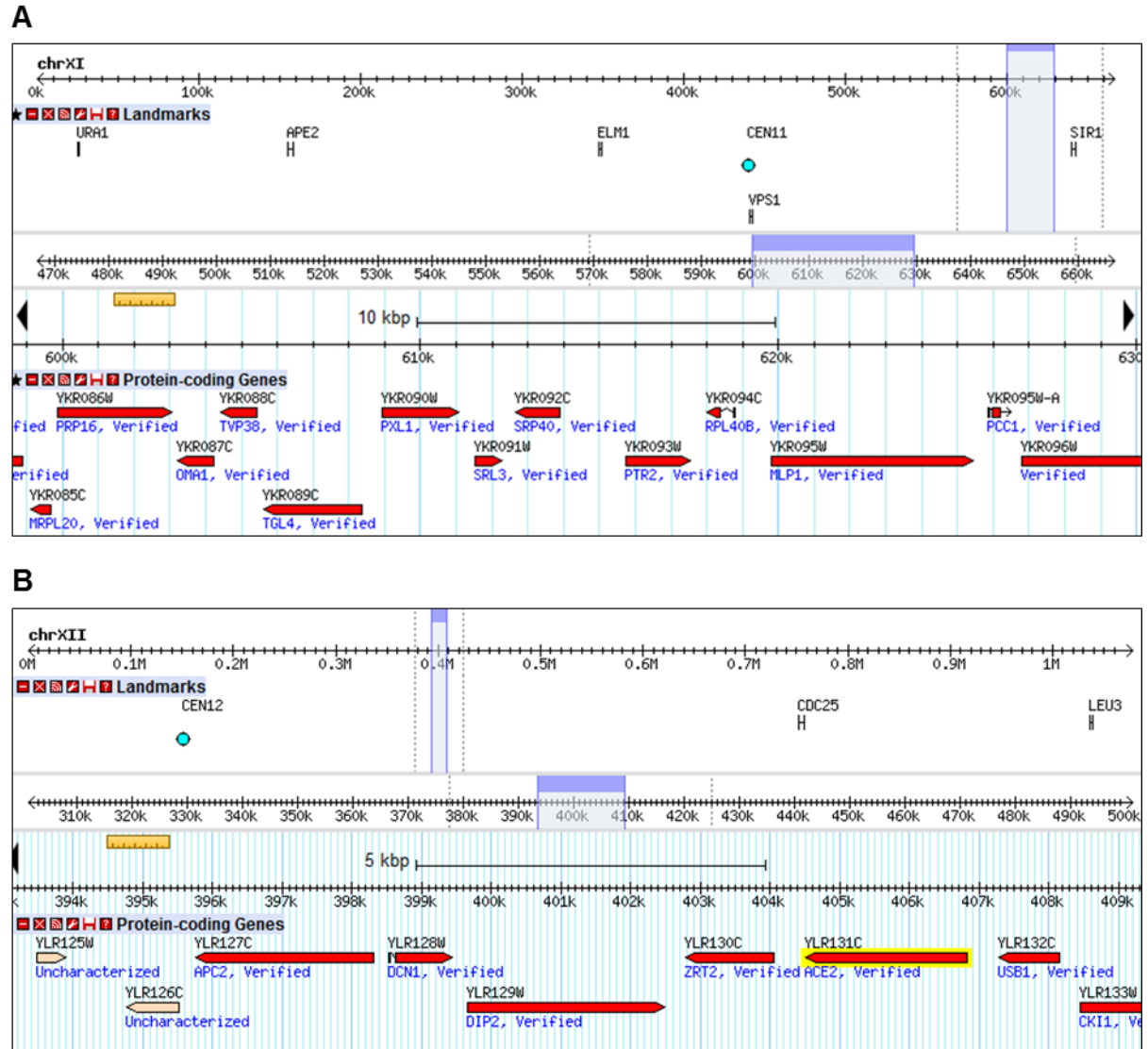
In addition, we induced sporulation of one of the spontaneous diploid isolates with altered colony morphology, JAY663. We then dissected tetrads, and examined the phenotypes of the haploid derivatives. None of the resulting haploids (~100) displayed the rough colony phenotype; they were all smooth. We then took twelve of these haploids derived from three four-spore tetrads (ARY51 through ARY62), determined their mating types, and conducted all possible mating combinations between them (Figure 3B). In this case, all 36 crosses resulted in rough colony diploids. This result was consistent with JAY663 being homozygous for the causal recessive mutant allele (*r/r*), and supported the hypothesis that LOH could be responsible for the sporadic appearance of the mutant phenotype in JAY270.



**Figure 3 - Cross between PE-2 haploids.** (A) Twelve haploids derived from the diploid JAY270 wild-type strain had their whole genomes sequenced. The cross showed a 1:4 phenotypic ratio, expected for a trait controlled by one recessive allele and the rough colony morphology phenotype only appeared in diploid cells. (B) The cross between haploids derived from three completed tetrads of the diploid JAY663 wild-type strain with spontaneous rough colonies generated a 36 population presenting rough colonies and again the phenotype only appeared in diploid cells.

## Identification of the mutation responsible for the rough colony phenotype

Based on the interpretation that the rough colony phenotype was determined by monogenic recessive inheritance of a diploid-dependent trait (Figure 3A), we divided the fully-sequenced JAY270-derived haploids in two groups according to their inferred genotype. Group 1 included the six haploids inferred to carry the mutant allele (*r*; JAY291, JAY292, JAY295, JAY296, JAY325, and JAY326); whereas group 2 was composed of the six haploids with the wild type allele (*R*; JAY289, JAY290, JAY293, JAY294, JAY327, and JAY328). We then compared the genome sequences of the twelve haploids to a map of high confidence heterozygous single nucleotide polymorphisms in the JAY270 diploid. This map contains more than eleven thousand heterozygous marker loci distributed across the genome (F. M. Duarte, T. M. Gurol, A. Rodrigues, *et al. in preparation*). We interrogated each of these loci searching for SNPs that co-segregated in all six individuals within group 1, and that conversely had the other allele co-segregating in all six individuals within group 2. This analysis identified only two candidate regions that fit the strict co-segregation criterion (Figure 4). The first region corresponded to approximately 30 Kb on the distal portion of the right arm of chromosome XI, including thirteen genes from *MRPL20* to *YKR096W*. The second candidate region spanned about 15 Kb on the right arm of chromosome XII, located about 50 Kb centromere proximal of the repeat cluster encoding the ribosomal RNA genes (rDNA). This smaller region contained the nine genes from *YLR125W* to *CKI1*.



**Figure 4 - Candidate regions found in the genetic mapping.** We found two candidate regions that co-segregated in the group of six spores with one genotype JAY291, JAY292, JAY295, JAY296, JAY325, JAY326 as well as in the group containing six haploids with the opposite genotype JAY289, JAY290, JAY293, JAY294, JAY327 and JAY328. The first region (A) was located on chromosome XI, from genes YKR085C (*MRPL20*) to YKR096W of *Saccharomyces Genome Database* (SGD) and the second one (B) on chromosome XII from genes YLR125W to YLR133W (*CK11*). Then, we found that *ACE2* gene (highlighted in figure b), located on chromosome XII, was involved with the phenotype.

To identify the most promising candidates among the 22 genes present in these two regions, we analyzed their respective functional annotations available through the *Saccharomyces Genome Database*. This analysis immediately revealed a gene located in

the chromosome XII candidate region, *ACE2*, which encodes a transcription factor involved in the mother-daughter cell separation process. In cells lacking Ace2p, the daughter cell remains attached to the mother cell wall at the bud neck, resulting in the accumulation of multicellular clusters. Importantly, *ace2* mutant strains display a rough colony morphology. In some strain backgrounds, this phenotype is observed only in diploids, but can also appear in certain haploid backgrounds (Dohrmann *et al.* 1992; Lord *et al.* 2002; Voth *et al.* 2005).

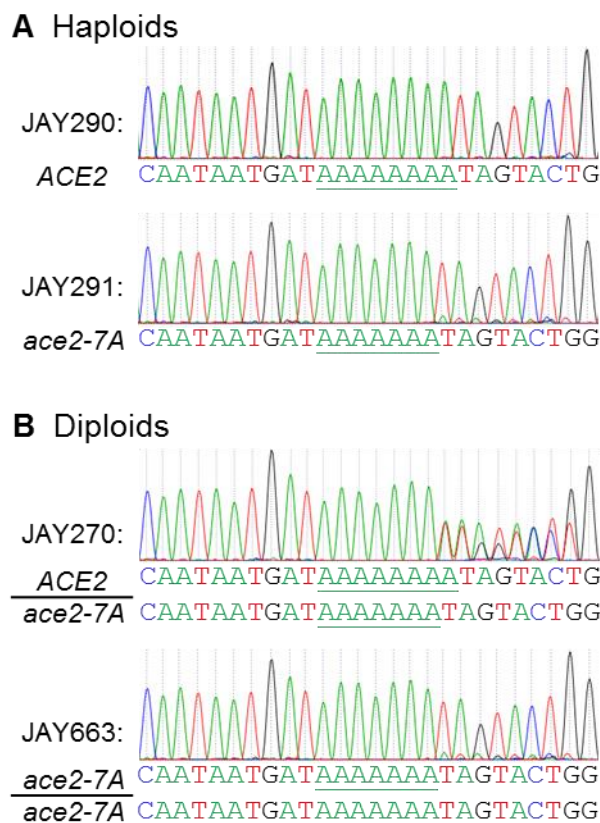
We inspected the sequence of the *ACE2* gene in the JAY291 genome, and compared it to the sequence in the S288c reference. Only one difference was identified between the two strains. JAY291 has a single nucleotide deletion within a homopolymer Adenine run resulting in a -1 frameshift mutation and a stop codon shortly thereafter. The wild type *ACE2* allele in S288c has eight A nucleotides in this homopolymer run, while the mutant allele in JAY291 has seven A nucleotides. Hence, we named the defective allele *ace2-7A*.

### **Analysis of the genomic region containing *ACE2* in spontaneous JAY270-derived rough colony isolates**

After finding that *ACE2* was the gene responsible for the rough colony morphology in crosses between JAY270 spores (Figure 3A), we revisited the five spontaneous rough colony isolates observed earlier in this study. We designed oligonucleotides flanking the adenine homopolymer run in *ACE2*, PCR amplified this region from JAY270, from its haploid derivatives JAY290 and JAY291, and from the spontaneous rough colony isolates. The PCR products were Sanger sequenced to generate the chromatograms shown in



Figure 5. This analysis confirmed the presence of a run of 8 adenines in *ACE2* (JAY290) and 7 adenines in *ace2-7A* (JAY291). The chromatogram in the JAY270 heterozygote diploid was consistent with a mixture of *ACE2* and *ace2-7A* DNA templates being present in the sequencing reaction: single nucleotide peaks were observed at positions primer-proximal to the homopolymer run, and out-of-register double peaks were seen downstream of the seventh adenine nucleotide.



**Figure 5 - Chromatograms of *ACE2* region that contains wild-type or mutant alleles in PE-2 strains.** Haploid strains JAY290 and JAY291 have the wild-type (*ACE2*) and the mutant (*ace2-7A*) alleles respectively (A). The diploid JAY270 is heterozygous for *ACE2*, presenting one wild-type functional allele and the other one has a frameshift mutation (*ACE2/ace2-7A*). Wild-type diploid strain with spontaneous rough colonies JAY663 is an example of homozygosity for the mutation (*ace2-7A/ace2-7A*) (B).

The chromatograms for all five rough colony isolates showed the presence of the *ace2-7A* frameshift mutation and a lack of the *ACE2* allele (Figure 5, and data not shown).

These results indicated that the rough colony isolates were homozygous for *ace2-7A*, and were consistent with the hypothesis that the high frequency of observed spontaneous rough colonies in JAY270 derivatives might be caused by mitotic recombination leading to LOH, specifically on the right arm of chromosome XII spanning the *ACE2* locus.

LOH is not typically a local mutation event: Interstitial tracts of homozygosity can span many kilobases, and even longer terminal tracts extend all the way to the telomere. Therefore, in addition being homozygous for *ace2-7A*, the rough colony isolates should also show homozygosity for flanking SNPs. We tested this model by determining the genotypes at chromosome XII heterozygous marker loci using either restriction length polymorphisms or Sanger DNA sequencing.

Analysis of the JAY663 isolate showed that, while it was homozygous for the *ace2-7A* mutation, it remained heterozygous at the sites proximal and distal to the *ACE2* locus, which are separated by a distance of 8.7 Kb. A mitotic gene conversion tract limited to the region between these SNPs could explain this result. The median size of mitotic gene conversion tracts not-associated with reciprocal crossovers in *S. cerevisiae* has recently been estimated to be 4.9 Kb (Y. Yin and T. D. Petes, personal communication). Alternatively, a *de novo* -1 contraction mutation in the homopolymer run of the *ACE2* allele could also account the for the JAY663 genotype. We were unable to distinguish between the LOH vs. *de novo* -1 contraction models because there were no additional heterozygous markers in this region that could be sequenced.

The four remaining isolates (JAY664, JAY665, JAY912, and JAY913) were homozygous for regions flanking the *ACE2* locus, suggesting the switch of colony morphology in PE-2/JAY270 is due to loss-of-heterozygosity in the *ACE2* region.

## DISCUSSION

### **PE-2 colony morphology transition is due to a mutation in *ACE2*, a gene involved in the cell separation process**

The spontaneous JAY270 colony variation found in this study is result of a cell separation defect. We found different isolates derivative from JAY270 carrying copy-neutral homozygosity to a frameshift mutation in the *ACE2* gene.

Ace2p is a transcription factor that regulates the expression of several genes. In *Saccharomyces cerevisiae*, at least 20 of these genes can be found (Doolin *et al.* 2001), some of them involved in the cell separation process, such as *SCW11* and *EGN1*, responsible for transcription of glucanases (Bidlemaier *et al.* 2001; Baladrón *et al.* 2002) and *CTS1*, encoding a chitinase (Kuranda and Robbins 1991). These enzymes operate in the digestion process of the chitin-rich septum existing between mother and daughter cells in the end of mitosis minutes after it is built and just after the cytokinesis is completed (Bidlemaier *et al.* 2001; Baladrón *et al.* 2002; Cabib 2004; Brace *et al.* 2011).

Ace2p contains a region in the DNA-binding domain that is conserved between representatives of different groups of Fungi, including *S. cerevisiae*. It was proposed that Ace2p had evolved in parallel with the evolution of the genomes to control other processes and the gain of cell separation function appeared at different and independent

times during the evolution of fungi, as it was needed. The novel target genes varied according to the composition and structure of the septa in different lineages (Balazs *et al.* 2012).

**The rough morphology only appears in diploid cells and mutations in other genes of the pathway can also lead to the same phenotype as *ace2***

In our study, we observed the rough colony morphology only in diploid homozygous *ace2* mutant cells. According to Voth and collaborators (2005) the colony morphology also depends on the genotype of another gene, *BUD4*, required for normal budding patterns. Normal haploid cells grow in an axial budding pattern, while the bipolar program is typical for diploids. Mutations in *BUD4* can switch the axial to a bipolar pattern in haploids, but have no effect in diploids. In the same study, *ace2* mutants with roughened colony morphology were only seen in diploid cells or combined with a *bud4* mutation in haploids (Voth *et al.* 2005).

The *ACE2* gene is regulated by a set of proteins, including Cbk1p kinase, Mob2p, Tao3p, Hym1p, Kic1p, which are all members of a signaling network known as RAM, meaning *regulation of Ace2p activity and cellular morphogenesis* (Doolin *et al.* 2001; Nelson *et al.* 2003; Kurischko *et al.* 2005). The activation of Ace2p depends on Cbk1p and signals from FEAR (Cdc *fourteen early anaphase release*) and MEN (*mitotic exit network*) pathways regulate *CBK1* and *ACE2* to act on the perfect timing for septum destruction and subsequent cell separation (Racki *et al.* 2000; Bidlingmaier *et al.* 2001; Weiss *et al.* 2002; Brace *et al.* 2011). Mutations in the RAM genes lead to lethality in some genetic backgrounds associated with a polymorphism in the *SSD1* gene, being essential when the

*SSD1* is functional (Nelson *et al.* 2003). However, when the mutation is viable, the mutants show a cell separation defect and altered colony morphology, similar to the one observed in *ace2* mutants, where the cells lose the typical hemispherical and smooth colony phenotype, but exhibit a roughened surface and scalloped edges instead, as observed in *cbk1*, *mob2*, *hym1*, and *kic1* mutants (Voth *et al.* 2005).

### **The main mechanism leading to PE-2 colony variation is Loss-of-Heterozygosity (LOH)**

We found that PE-2 is heterozygous to *ace2* mutation and PE-2 strains with spontaneous roughened colonies found in the laboratory are in the majority homozygous to this mutation. The high frequency in which colony morphology variation was observed indicated that it was not likely due to *de novo* nucleotide mutation, but instead mitotic recombination leading to loss-of-heterozygosity. We confirmed this hypothesis by sequencing regions proximal and distal to *ACE2* in both PE-2 strains with smooth and rough colonies. Our results showed that JAY270 is heterozygous at selected regions, while JAY270-derivative spontaneous mutants were homozygous, indicating that the main mechanism leading to PE-2 colony variation found in this study is LOH.

## **CONCLUSIONS**

Our findings showed that the transition from typical smooth colonies of PE-2/JAY270 bioethanol strain to roughened surfaces and high sedimentation phenotype resulted from a defect in the cell separation process. In this study we found that strains with the rough morphology were homozygous to *ace2-7A* frameshift mutation, a gene

involved in the mother-daughter cell separation and observed that the main mechanism leading to this phenotype was loss-of-heterozygosity.

## REFERENCES

- Amorim, H. V, M. L. Lopes, J. V. de Castro Oliveira, M. S. Buckeridge, and G. H. Goldman, 2011 Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 1267–1275.
- Argueso, J. L., M. F. Carazzolle, P. A. Mieczkowski, F. M. Duarte, O. V. C. Netto *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19: 2258–70.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.*, 1998 *Current Protocols in Molecular Biology*. John Wiley & Sons Inc., New York.
- Baladrón, V., S. Ufano, E. Dueñas, A. B. Martín-Cuadrado, F. del Rey *et al.*, 2002 Eng1p , an endo-1,3- $\beta$ -glucanase localized at the daughter side of the septum , is involved in cell separation in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 1: 774–786.
- Balazs, A., G. Batta, I. Miklos, L. Acs-Szabo, C. R. Vazquez de Aldana *et al.*, 2012 Conserved regulators of the cell separation process in *Schizosaccharomyces*. *Fungal Genet. Biol.* 49: 235–49.
- Basso, L. C., H. V de Amorim, A. J. de Oliveira, and M. L. Lopes, 2008 Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res.* 8: 1155–63.
- Bidlingmaier, S., E. L. Weiss, C. Seidel, D. G. Drubin, and M. Snyder, 2001 The Cbk1p pathway is important for polarized cell growth and cell separation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 21: 2449–2462.
- Boeke, J. D., F. Lacroute, and G. R. Fink, 1984 A positive selection for mutants lacking orotidine-5 ' -phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* 197: 345–346.
- Brace, J., J. Hsu, and E. L. Weiss, 2011 Mitotic exit control of the *Saccharomyces cerevisiae* Ndr/LATS kinase Cbk1 regulates daughter cell separation after cytokinesis. *Mol. Cell. Biol.* 31: 721–35.
- Cabib, E., 2004 The septation apparatus, a chitin-requiring machine in budding yeast. *Arch. Biochem. Biophys.* 426: 201–7.

- Dohrmann, P. R., G. Butler, K. Tamai, S. Dorland, J. R. Greene *et al.*, 1992 Parallel pathways of gene regulation: homologous regulators *SWI5* and *ACE2* differentially control transcription of *HO* and chitinase. *Genes & Dev.* 6: 93–104.
- Doolin, M. T., a L. Johnson, L. H. Johnston, and G. Butler, 2001 Overlapping and distinct roles of the duplicated yeast transcription factors Ace2p and Swi5p. *Mol. Microbiol.* 40: 422–32.
- Halme, A., S. Bumgarner, C. Styles, and G. R. Fink, 2004 Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. *Cell* 116: 405–15.
- Honigberg, S. M., 2011 Cell signals, cell contacts, and the organization of yeast communities. *Eukaryot. Cell* 10: 466–473.
- Huxley, C., G. ED, and I. Dunham, 1990 Rapid assessment of *S. cerevisiae* mating type by PCR. *Trends Genet.* 6: 236.
- Kuranda, M. J., and P. W. Robbins, 1991 Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 266: 19758–67.
- Kurischko, C., G. Weiss, M. Ottey, and F. C. Luca, 2005 A role for the *Saccharomyces cerevisiae* regulation of Ace2 and polarized morphogenesis signaling network in cell integrity. *Genetics* 171: 443–55.
- Lord, M., T. Chen, A. Fujita, and J. Chant, 2002 Analysis of budding patterns. *Methods Enzymol.* 350: 131–141.
- Van Mulders, S. E., E. Christianen, S. M. G. Saerens, L. Daenen, P. J. Verbelen *et al.*, 2009 Phenotypic diversity of Flo protein family-mediated adhesion in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 9: 178–90.
- Nelson, B., C. Kurischko, J. Horecka, M. Mody, P. Nair *et al.*, 2003 RAM: a conserved signaling network that regulates Ace2p transcriptional activity and polarized morphogenesis. *Mol. Biol. Cell* 14: 3782–3803.
- Racki, W. J., A. M. Bécam, F. Nasr, and C. J. Herbert, 2000 Cbk1p, a protein similar to the human myotonic dystrophy kinase, is essential for normal morphogenesis in *Saccharomyces cerevisiae*. *EMBO J.* 19: 4524–32.
- Soares, E. V, 2011 Flocculation in *Saccharomyces cerevisiae*: a review. *J. Appl. Microbiol.* 110: 1–18.
- Verstrepen, K. J., and F. M. Klis, 2006 Flocculation, adhesion and biofilm formation in yeasts. *Mol. Microbiol.* 60: 5–15.

- Voth, W. P., A. E. Olsen, M. Sbia, K. H. Freedman, and D. J. Stillman, 2005 *ACE2*, *CBK1*, and *BUD4* in budding and cell separation. *Eukaryot. Cell* 4: 1018–1028.
- Weiss, E. L., C. Kurischko, C. Zhang, K. Shokat, D. G. Drubin *et al.*, 2002 The *Saccharomyces cerevisiae* Mob2p-Cbk1p kinase complex promotes polarized growth and acts with the mitotic exit network to facilitate daughter cell-specific localization of Ace2p transcription factor. *J. Cell. Biol.* 158: 885–900.
- Wheals, A. E., L. C. Basso, D. M. Alves, and H. V Amorim, 1999 Fuel ethanol after 25 years. *Trends Biotechnol.* 17: 482–7.



## **Capítulo 3**

**Manipulação genética de linhagem PE-2 para reparo de defeito  
identificado em gene envolvido com variação de morfologia de  
colônia**



## **Manipulação genética de linhagem PE-2 para reparo de defeito identificado em gene envolvido com variação de morfologia de colônia**

### **INTRODUÇÃO**

No capítulo anterior, foi apresentado um estudo onde encontramos mutantes espontâneos de PE-2 com morfologia de colônia alterada, exibindo superfícies rugosas, bordas irregulares e um fenótipo que remete ao de floculação em meio líquido.

Em outros estudos, foi observada a presença de linhagens rugosas, incluindo PE-2, durante fermentação industrial. As linhagens rugosas são frequentemente associadas com crescimento em pseudohifas e alta taxa de sedimentação, causando problemas similares aos de linhagens floculantes em um ambiente industrial. A floculação prejudica o processo de produção de etanol, pois reduz o contato da levedura com o substrato, reduzindo o rendimento fermentativo e resultando em açúcar residual após a fermentação, além de dificultar o passo de centrifugação e causar entupimento de tubulações (Basso *et al.* 2008).

Identificamos o gene *ACE2* como responsável pela variação de morfologia de colônia em PE-2/JAY270 e vimos que essa linhagem é heterozigota para um defeito nesse gene (*ACE2/ace2-7A*). Nossas descobertas mostraram que esse defeito pode resultar no aparecimento da morfologia de colônia rugosa por perda de heterozigosidade (Capítulo 3). Além disso, o gene *ACE2* aparece como um gene candidato associado com floculação em nossos estudos de mapeamento genômico (Capítulo 2), provavelmente como um gene de efeito menor, mas atuante como modificador do fenótipo. Com o intuito de se evitar

problemas industriais futuros, nesse capítulo apresentamos o desenvolvimento de uma linhagem PE-2 com o defeito em *ACE2* reparado, através de modificações genéticas para que se tornasse homozigota para *locus* selvagem desse gene (*ACE2/ACE2*).

## MATERIAIS E MÉTODOS

### Meios de cultura, condições de crescimento, linhagens e técnicas microbiológicas/moleculares

YPD, Luria-Bertani (LB), meio de esporulação e meio mínimo foram preparados conforme descrições de protocolos padrões (Ausubel *et al.* 1998). *Escherichia coli* foi incubada em estufa à temperatura de 37°C, em meio LB suplementado com 100 µg/ml de ampicilina. Linhagens de leveduras foram incubadas em estufa à temperatura de 30°C, em meio mínimo ou YPD. Foi utilizado suplemento sintético sem uracila (*Yeast Synthetic Dropout Media Supplement without Uracil* - Sigma-Aldrich) para preparo de meio mínimo de seleção de transformantes que integraram o plasmídeo pAR1. O composto 5-FOA (5-fluoro-orotic acid) foi suplementado em meio mínimo a uma concentração final de 1 g/l (Boeke *et al.* 1984) para seleção de leveduras que perderam pAR1. As linhagens de leveduras utilizadas nesse capítulo estão descritas na Tabela 1. A linhagem *E. coli* (*α-Select Chemically Competent Cells*, Bioline) foi utilizada para preparo do plasmídeo e a transformação foi realizada de acordo com especificações do fabricante. Para dissecação de tétrades, a linhagem ARY103 foi plaqueada em meio de esporulação e incubada em estufa à temperatura de 30°C, durante três dias. Após esporulação, as tétrades foram dissecadas através de micromanipulação. Técnicas de extração de DNA, transformação de leveduras

pelo método acetato de lítio, esporulação, dissecação de tétrades e cruzamento de linhagens estão descritas em protocolos padrões (Rose *et al.* 1990; Ausubel *et al.* 1998). Os *mating types* das linhagens haploides ou diploides, *MAT $\alpha$* , *MAT $\alpha$*  or *MAT $\alpha$ / $\alpha$*  foram confirmados por PCR, conforme técnica descrita por (Huxley *et al.* 1990). Os *primers* utilizados em PCRs realizados nesse trabalho são apresentados na Tabela 2. Os sequenciamentos de fragmentos de PCR ou plasmídeo construído foram realizados em equipamento *ABI 3130xL Genetic Analyzer* (Applied Biosystems), pela *Proteomics and Metabolomics Facility*, localizada na *Colorado State University*, Fort Collins, CO, EUA. A análise das sequências foi realizada com utilização do software *CLC Genomics Workbench*.

Tabela 1 - Linhagens de levedura utilizadas nesse estudo

Genótipos:		Descrições:	Referências:
<b>Diploides:</b>			
JAY270:	$\frac{MAT\alpha}{MAT\alpha}$ $\frac{ACE2}{ace2}$	Isolado selvagem de PE-2	Argueso <i>et al.</i> 2009
FGY050:	$\frac{MAT\alpha}{MAT\alpha}$ $\frac{ACE2}{ace2}$ $\frac{ura3\Delta}{ura3\Delta}$	Derivada de JAY270; auxotrófica para uracila	Galzerani 2010
ARY103, ARY104:	$\frac{MAT\alpha}{MAT\alpha}$ $\frac{ACE2}{ACE2}$	Mutação em <i>ACE2</i> reparada no genoma de FGY050	Esse estudo
<b>Haploides:</b>			
ARY105, ARY106, ARY107, ARY108:	Olhar figura 7	Tétrade de ARY103	Esse estudo
ARY109, ARY110, ARY111, ARY112:	Olhar figura 7	Tétrade de ARY103	Esse estudo

ARY113, ARY114, Olhar figura 7  
ARY115, ARY116:

Tétrade de ARY103

Esse estudo

**Tabela 2 - Primers utilizados nesse trabalho**

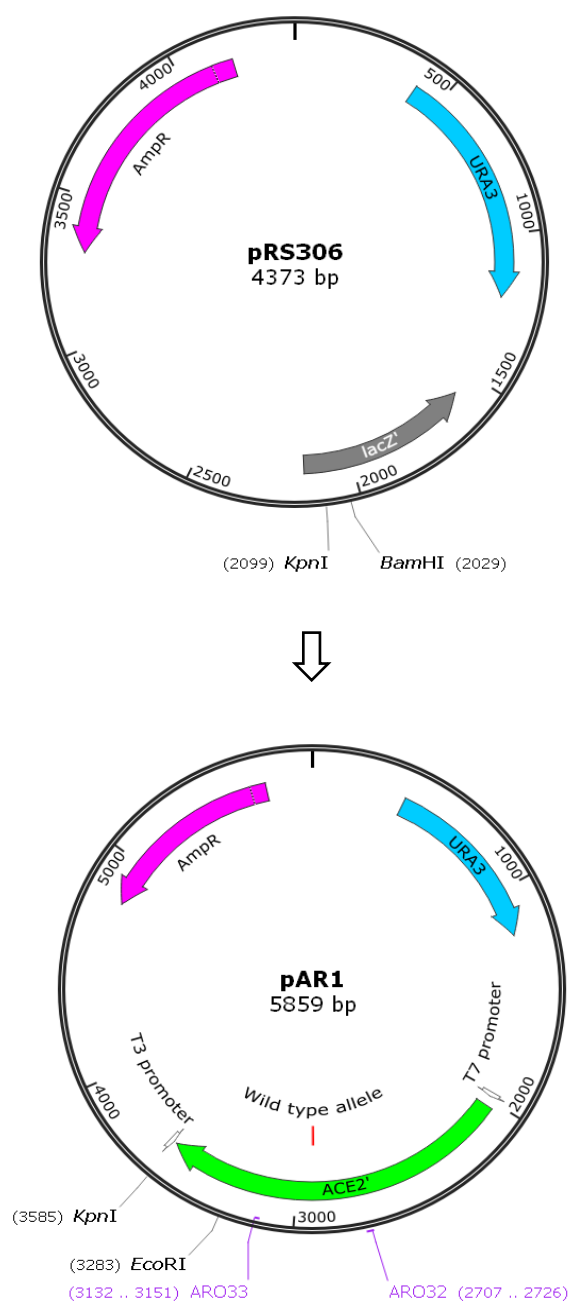
<b>Nomes:</b>	<b>Sequências 5' → 3'</b>	<b>Descrições:</b>
ARO30	ACACTCAAGATGAGGAGTAT	Forward - construção de pAR1
ARO31	TCGCATGAATACGGTCTATC	Reverse - construção de pAR1
ARO32	TACTTCAATAAATGGTTCAC	Forward - Homologia interna ao gene <i>ACE2</i>
ARO33	TGATATTGTCGAGACCGTGG	Reverse - Homologia interna ao gene <i>ACE2</i>
ARO38	TAGCACAAGATATGAACT	Forward - região proximal <i>ACE2</i>
ARO39	TATATCATCTACAATGAC	Reverse - região proximal <i>ACE2</i>
ARO70	ATACTGGAGAGTGTTGGG	Forward - região distal <i>ACE2</i>
ARO71	TCTAGCGACCAAATTGCC	Reverse - região distal <i>ACE2</i>

## Construção de plasmídeo

O plasmídeo pRS306 (Sikorski and Hieter 1989), vetor integrativo de levedura portador de marcador auxotrófico *URA3*, foi utilizado como hospedeiro para a construção do plasmídeo. A construção foi realizada em dois passos. Primeiramente, pRS306 foi duplamente digerido pelas enzimas de restrição *BamHI* e *KpnI* (New England Biolabs). O fragmento contendo o gene *URA3* foi purificado de gel de agarose 0,8% (*GeneJET Gel Extraction Kit* - Thermo Scientific) e reservado. No segundo passo, sequências parciais do alelo selvagem de *ACE2* foram amplificadas por PCR utilizando-se DNA genômico de JAY289 e os *primers* ARO30 e ARO31. Com utilização da enzima de alta fidelidade *Velocity DNA Polymerase* (Bioline), a amplificação foi iniciada à temperatura de 98°C durante 50 segundos e então realizada em 33 ciclos de 98°C durante 10 segundos, 55°C durante 18 segundos e 72°C durante 2 minutos, com um passo final de 4 min à temperatura de 72°C. O amplificado foi purificado (*GeneJET PCR Purification Kit* - Thermo Scientific) e então

digerido com as enzimas de restrição *Bgl*III e *Kpn*I (New England Biolabs). O fragmento contendo sequência parcial do gene *ACE2* foi purificado de gel de agarose 1% (*GeneJET Gel Extraction Kit* - Thermo Scientific) e então utilizado para reação de ligação com o fragmento de pRS306, reservado anteriormente.

Após clonagem em *E. coli*, o plasmídeo foi isolado com utilização de kit comercial (*GeneJET Plasmid Miniprep Kit* - Thermo Scientific). O plasmídeo obtido, contendo sequência parcial do tipo selvagem de *ACE2* e marcador auxotrófico *URA3* foi designado como pAR1 e sequenciado com *primers* universais T3 e T7 antes de transformação em levedura.



**Figura 1 - Construção de plasmídeo pAR1 a partir de pRS306.** O plasmídeo pAR1 foi construído utilizando-se como base o plasmídeo pRS306 (Sikorski and Hieter 1989). pAR1 contém sequências parciais do alelo selvagem de ACE2 e marcador auxotrófico URA3 e foi utilizado para manipulação genética no genoma da linhagem PE-2/JAY270. Figura gerada através do programa SnapGene Viewer 1.5.2.



## Construção de linhagem PE-2 homozigota para o *locus* selvagem de *ACE2*

A linhagem de levedura utilizada para construção de PE-2 homozigota para o *locus* selvagem de *ACE2* foi a FGY050, auxotrófica para uracila (Galzerani 2010), permitindo assim manipulações genéticas. A construção da linhagem foi realizada sem deixar qualquer marca no final do processo em seu genoma, o que é importante para a aplicação e uso industrial da linhagem em processos fermentativos.

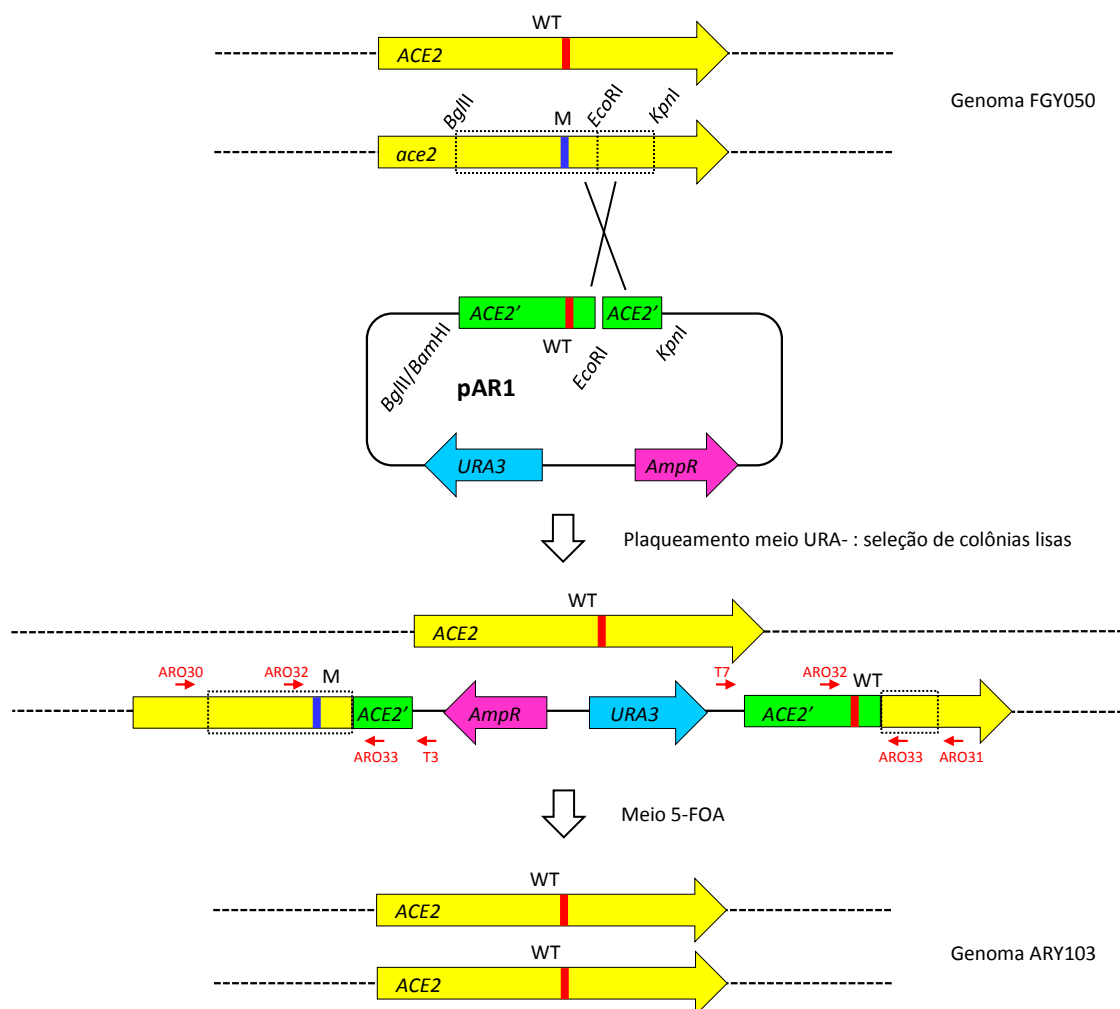
O plasmídeo pAR1 foi digerido com *EcoRI* (New England Biolabs) e então o fragmento linear integrativo resultante foi inserido no genoma de FGY050. Após a transformação, as células de levedura foram plaqueadas em meio mínimo com ausência de suplemento aminoácido uracila para seleção de transformantes *URA3*. Dois candidatos que cresceram nesse meio foram selecionados e a inserção de pAR1 foi confirmada por sequenciamento de fragmento de PCR realizado como descrito a seguir.

Para o sequenciamento, duas diferentes amplificações foram realizadas em cada candidato. Os *primers* externos ARO30, ARO31 e universais internos ao plasmídeo pAR1 T3 e T7 foram utilizados para essas amplificações (Figura 2). A primeira reação ocorreu utilizando-se os *primers* ARO31 e T7, *Velocity DNA Polymerase* (Bioline) e foi iniciada 98°C durante 50 segundo e então realizada com 35 ciclos de 98°C durante 10 segundos, 55°C durante 18 segundos e 72°C durante 3 minutos, com uma extensão final de 4 minutos à 72°C.

O segundo foi realizado da mesma maneira que o primeiro, mas com utilização dos *primers* ARO30 e T3. O fragmento produto gerado no primeiro caso foi sequenciado em três reações independentes com os *primers* ARO31, ARO32 e T7. O produto amplificado

do segundo PCR também foi sequenciado em três reações diferentes com os *primers* ARO30, T3 e ARO32.

Após confirmação, quinze colônias purificadas de cada transformante foram semeadas em meio sólido contendo o composto 5-FOA, que é seletivo para linhagens *ura3* (Boeke *et al.* 1984). Cinco candidatos selecionados que cresceram em placas 5-FOA foram sequenciados para o gene *ACE2* para confirmar a perda de pAR1 e reparo da mutação em frameshift *ace2-7A*. As amplificações foram realizadas com *MyTaq DNA Polymerase* (Bioline), com os *primers* ARO30 e ARO33 e sequenciadas com ARO33. Após confirmação, as duas linhagens finais foram nomeadas ARY103 e ARY104. A construção da linhagem está esquematizada na Figura 2.

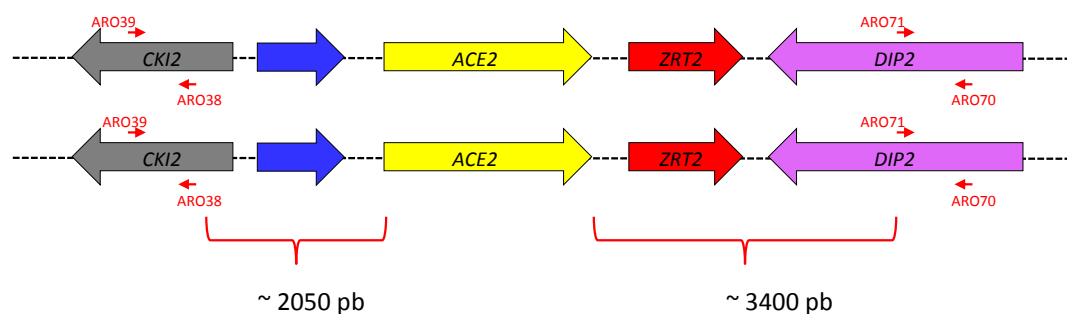


**Figura 2 - Esquematização de estratégia utilizada para manipulação genética da linhagem PE-2/JAY270.** O plasmídeo *pAR1*, que contém sequências parciais do alelo selvagem de *ACE2* (WT-vermelho) e marcador *URA3* foi utilizado para reparo da mutação *ace2-7A* (M-azul) no genoma de FGY050. Os transformantes *URA3* de colônias lisas foram selecionados em meio mínimo com deficiência do aminoácido uracila e linhagens *ura3Δ* contra-selecionadas em meio contendo o composto 5-FOA. A linhagem resultante homocigota para alelo selvagem (WT) de *ACE2* foi nomeada ARY103.

As linhagens finais ARY103 e ARY104 foram sequenciadas também em regiões proximal e distal à *ACE2* para checar heterozigosidade.

A região proximal foi amplificada com os *primers* ARO38 e ARO39 e a região distal com os *primers* ARO70 e ARO71, ambas sequenciadas com ARO38 e ARO70

respectivamente (Figura 3). Também foi sequenciado o fragmento gerado por PCR com *primers* ARO32 e ARO33 para as linhagens ARY103 e ARY104 com ARO32.



**Figura 3 - Regiões próximas ao gene *ACE2* sequenciadas para checar heteroziguidade de ARY103 e ARY104.** Após construção de PE-2 homozigota para alelo selvagem de *ACE2*, checamos se a manipulação manteve a heteroziguidade natural da linhagem através de PCR, nas regiões proximal (*primers* ARO38 e ARO39) e distal (*primers* ARO70 e ARO71), localizadas respectivamente a aproximadamente 2050 pb e 3400 pb do gene *ACE2*.

## RESULTADOS

### Estratégia para reposição alélica utilizada para construção de linhagem PE-2 homozigota para *locus* tipo selvagem de *ACE2*

Como a mudança de morfologia de colônia de lisa para rugosa é indesejada para a indústria produtora de etanol, nós modificamos o genoma da PE-2/JAY270 para homozigota para *locus* selvagem de *ACE2* e consequentemente incapaz de se tornar rugosa por LOH. Com esse objetivo, construímos um plasmídeo com marcador *URA3* e fragmento tipo selvagem de *ACE2* e inserimos em uma linhagem derivada de JAY270, auxotrófica para uracila (FGY050). Nós então selecionamos transformantes *URA3* com colônias lisas, já que a inserção do plasmídeo no alelo selvagem da linhagem heterozigota FGY050 leva a um diploide com dois alelos não funcionais de *ACE2* e, consequentemente, fenótipo rugoso. Subsequentemente, semeamos transformantes lisos em placas 5-FOA,

que é tóxica para linhagens *URA3* e então contra selecionamos células que perderam sequências do plasmídeo. A remoção de pAR1 ocorreu após evento de recombinação entre sequências homólogas (repetidas) de *ACE2* que agora flanqueiam o marcador *URA3* (após integração de pAR1 no genoma de FGY050), mantendo apenas uma única cópia da repetição de DNA e assim, a linhagem resultante passou a apresentar duas cópias do alelo selvagem de *ACE2* (Figura 2).

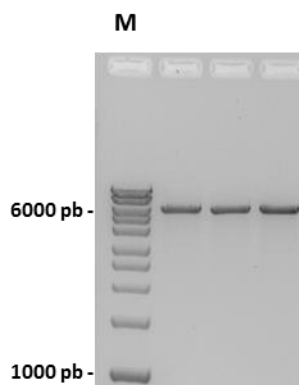
A estratégia nos permitiu reparar o mutante para alelo tipo selvagem de *ACE2* no genoma de FGY050, sem deixar qualquer tipo de marcador ou sequências do vetor, importante para manter as propriedades industrialmente desejadas da linhagem.

### **Construção do plasmídeo pAR1**

Primeiramente, nós construímos o plasmídeo pAR1, que contém o marcador *URA3* e o fragmento tipo selvagem de *ACE2*, baseado no plasmídeo pRS306. A digestão de pRS306 *BamHI* e *KpnI* produziu um fragmento de 4,3 kb, contendo o marcador *URA3*. Por outro lado, a digestão do fragmento tipo selvagem de *ACE2* amplificado de JAY289 de 1,8 kb com *BglII* e *KpnI* gerou um produto de 1,5 kb, inserido nos sítios de restrição do fragmento de 4,3 kb *BamHI-KpnI*-pRS306. O plasmídeo resultante de 5,8 kb foi sequenciado para confirmar sua construção correta e garantir que nenhuma mutação apareceu no fragmento gênico de *ACE2* inserido, durante o processo de construção.

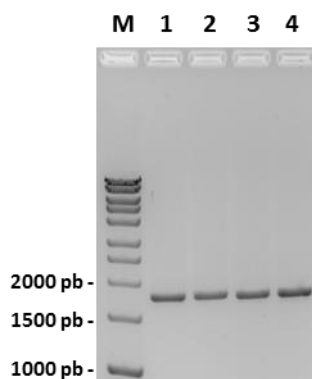
### **Reparo do gene *ACE2* em uma linhagem PE-2**

Após confirmação da correta construção de pAR1 pelo sequenciamento, reparamos o gene *ACE2* na linhagem diploide PE-2/FGY050, utilizando o fragmento linear integrativo de 5,8 kb gerado pela digestão de pAR1 com *EcoRI* (Figura 4).



**Figura 4 - Fragmento gerado pela digestão de pAR1 com a enzima de restrição *EcoRI*.** Gel de agarose 1%, contendo amostras de reação de digestão do plasmídeo pAR1 com enzima *EcoRI*. O fragmento resultante de 5,8 kb foi utilizado durante construção da linhagem PE-2/JAY270 homóloga para alelo selvagem de *ACE2*. Marcador de peso molecular “HyperLadder I” (1 kb) – Bioline (M).

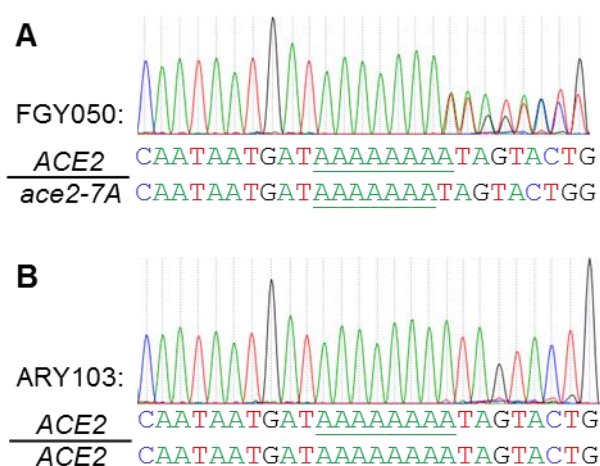
Após transformação de FGY050 com o fragmento acima descrito, a integração homóloga de pAR1 foi confirmada por sequenciamento (descrito em materiais e métodos) dos produtos amplificados em duas reações de PCR, uma com utilização dos *primers* ARO31 (externo à inserção) e T7 (universal, específico de pAR1) e outra com os *primers* ARO30 (externo à inserção) e T3 (universal, específico de pAR1), ambos contendo aproximadamente 1,7 kb (Figura 5).



**Figura 5 - Confirmação da integração homóloga de fragmento de pAR1 por PCR, seguido de sequenciamento.** Gel de agarose 1% contendo amostras das duas reações de PCRs realizadas em transformantes candidatos à integração homóloga de pAR1. As amplificações ocorreram com os *primers* ARO31 e T7 (1 e 2); ARO30 e T3 (3 e 4); e ambas geraram fragmentos com aproximadamente 1,7 kb. Os produtos das reações foram sequenciados confirmando a correta integração homóloga do fragmento de pAR1. Marcador de peso molecular “HyperLadder I” (1 kb) – Bioline (M).

Após análise dos dados gerados pelo sequenciamento dos fragmentos, foi confirmada a inserção de pAR1 no genoma de FGY050 e logo em seguida, os transformantes *URA3* foram plaqueados em meio contendo o composto 5-FOA. Após crescimento das colônias *ura3Δ*, realizamos PCR para cinco candidatos selecionados com os *primers* ARO30 e ARO33. A perda de pAR1 e o reparo da mutação em frameshift *ace2-7A* foram confirmados pelo sequenciamento dos fragmentos produtos do PCR.

Somente um candidato apresentou o gene *ACE2* ainda heterozigoto para a mutação e para os outros quatro candidatos o reparo foi confirmado, mostrando que agora são homozigotos para o alelo selvagem de *ACE2*. Duas linhagens que tiveram a confirmação pelo sequenciamento foram salvas como ARY103 e ARY104 e então estocadas (Figura 6).



**Figura 6 - Confirmação da construção da linhagem ARY103 por sequenciamento.** Confirmamos o reparo da mutação *ace2-7A*, realizado no genoma da linhagem PE-2 (FGY050), heterozigota para alelo mutante (*ACE2/ace2-7A*) (A), através de sequenciamento de fragmento de PCR. A linhagem resultante ARY103 (B) é homozigota para alelo selvagem de *ACE2* (*ACE2/ACE2*)(B).

Também sequenciamos as regiões proximal e distal a *ACE2* em ARY103 e ARY104 e confirmamos que essas linhagens ainda são heterozigotas em ambas as regiões,

mostrando que o alelo mutante não foi perdido por LOH e o reparo foi feito pela estratégia, mantendo a heterozigosidade natural do genoma de PE-2/JAY270.

Finalmente, nós isolamos doze esporos de ARY103, provenientes de três tétrades completas diferentes e realizamos todos os possíveis cruzamentos entre haploides *MATa* e *MATα*, resultando em 36 diploides *MATa/α* diferentes. Conforme esperado, os 36 diferentes diploides possuem colônias lisas com bordas regulares, confirmando novamente o reparo do defeito de separação celular causado pela mutação *ace2-7A* (Figura 7).

		<i>MATa</i>					
		ARY105	ARY107	ARY110	ARY112	ARY115	ARY116
<i>MATα</i>	ARY106	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ
	ARY108	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ
	ARY109	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ
	ARY111	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ
	ARY113	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ
	ARY114	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ

**Figura 7 - Cruzamento entre esporos de ARY103.** Todos os possíveis cruzamentos *MATa* e *MATα* entre doze haploides provenientes de três tétrades completas de ARY103 foram realizados. Os 36 indivíduos diploides *MATa/α* resultantes apresentaram colônias lisas (S – *smooth colonies*), confirmando novamente o genótipo homozigoto selvagem *ACE2/ACE2* de ARY103.

## DISCUSSÃO

Ao longo dos anos, o melhoramento de linhagens de leveduras para uso industrial foi foco de diversas pesquisas (Wheals *et al.* 1999), uma vez que estas precisam apresentar características que permitam a sua sobrevivência em ambiente altamente estressante, como o encontrado nas dornas de fermentação.



A linhagem industrial PE-2/FGY050 utilizada nesse estudo é auxotrófica para uracila e foi desenvolvida através da técnica conhecida como *Delitto Perfetto* (Storici and Resnick 2006) durante mestrado de um aluno do nosso grupo (Galzerani 2010).

Nossos estudos descritos no capítulo anterior (Capítulo 3) mostraram que a troca de colônias típicas da PE-2/JAY270 para superfícies rugosas acompanhadas de fenótipo parecido com o de flocculação é na verdade resultado de um defeito no processo de separação celular e o gene *ACE2* foi identificado como responsável por esse fenótipo. Além disso, *ACE2* deve ser um gene de efeito menor atuante sobre flocculação (Capítulo 2). Vimos também que colônias rugosas de linhagens derivadas de PE-2/JAY270 eram homozigotas para a mutação em frameshit *ace2-7A* e que as típicas colônias lisas de JAY270 eram heterozigotas para essa mutação. Dado que a troca de fenótipo em PE-2/JAY270 para morfologia de colônia rugosa é resultado de LOH, construímos uma linhagem homozigota para o alelo selvagem de *ACE2* (*ACE2/ACE2*), tornando-a incapaz de adquirir o fenótipo rugoso por LOH.

A linhagem PE-2/FGY050 foi utilizada como plataforma para a manipulação genética desenvolvida nesse capítulo, permitindo o reparo do defeito encontrado no gene *ACE2*, com manutenção das propriedades industriais originais desejáveis da linhagem e sem deixar nenhuma marca de seleção no final do processo.

## CONCLUSÕES

A linhagem PE-2/ARY103 desenvolvida nesse estudo passou a ser homozigota para alelo selvagem do gene *ACE2* e consequentemente incapaz de se tornar rugosa por perda de heterozigosidade (LOH).

A manipulação genética aplicada foi realizada com manutenção de todas as características originais da PE-2, sem deixar nenhuma marca transgene em seu genoma e a linhagem desenvolvida poderá ser utilizada para se evitar problemas industriais causados por linhagens rugosas dependentes de *ACE2*.

## REFERÊNCIAS

- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.*, 1998 *Current Protocols in Molecular Biology*. John Wiley & Sons Inc., New York.
- Basso, L. C., H. V de Amorim, A. J. de Oliveira, and M. L. Lopes, 2008 Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res.* 8: 1155–63.
- Boeke, J. D., F. Lacroute, and G. R. Fink, 1984 A positive selection for mutants lacking orotidine-5 ' -phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* 197: 345–346.
- Galzerani, F., 2010 Desenvolvimento de plataformas industriais: domesticação de leveduras para o melhoramento da fermentação alcoólica: Universidade Estadual de Campinas. Tese de Mestrado. 78p.
- Huxley, C., G. ED, and I. Dunham, 1990 Rapid assessment of *S. cerevisiae* mating type by PCR. *Trends Genet.* 6: 236.
- Rose, M. D., F. Winston, and P. Hieter, 1990 *Methods in yeast genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sikorski, R. S., and P. Hieter, 1989 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122: 19–27.
- Storici, F., and M. a Resnick, 2006 The *Delitto Perfetto* approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Method. Enzymol.* 409: 329–45.
- Wheals, A. E., L. C. Basso, D. M. Alves, and H. V Amorim, 1999 Fuel ethanol after 25 years. *Trends Biotechnol.* 17: 482–7.

## **Capítulo 4**

**Análise das consequências fenotípicas da perda de heterozigosidade (LOH) durante a fermentação de etanol**

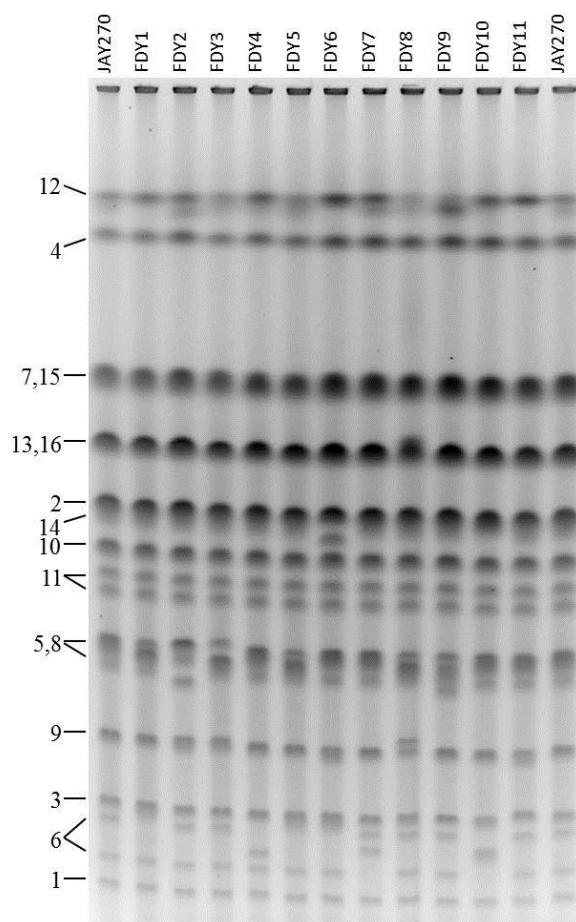


## **Análise das consequências fenotípicas da perda de heterozigosidade (LOH) durante a fermentação de etanol**

### **INTRODUÇÃO**

A instabilidade cariotípica durante o processo industrial de produção de etanol tem sido observada em linhagens industriais, incluindo a PE-2. Um estudo realizado pelo nosso grupo (Duarte 2010) determinou o mecanismo de recombinação genética responsável pela geração dos rearranjos cromossômicos na linhagem PE-2 durante o processo industrial de produção de etanol. Foram analisados os cariótipos de clones de leveduras isolados em um experimento de fermentação industrial contendo sessenta dias, iniciado com uma única célula da linhagem JAY270, isolado purificado do estoque comercial de PE-2 (Argueso *et al.* 2009). Onze desses isolados, apresentando no mínimo um polimorfismo de tamanho cromossomal relativamente a JAY270, foram analisados (Figura 1) e os dados indicaram que o mecanismo predominante para a variação cariotípica era recombinação homóloga alélica mitótica levando a perda de heterozigosidade (LOH).

Nesse capítulo, avaliamos fenotipicamente as onze linhagens LOH durante fermentação contínua de xarope de cana-de-açúcar em competição direta contra a linhagem parental altamente heterozigota JAY270 (Argueso *et al.* 2009), dando assim, continuidade aos estudos realizados por Duarte 2010. Nosso objetivo nessa etapa é investigar se os rearranjos cromossômicos e perda de heterozigosidade que ocorrem no genoma da PE-2 durante a produção de etanol estão relacionados com alguma alteração fenotípica benéfica à competitividade da linhagem.



**Figura 1 - Gel de agarose de PFGE (*Pulsed-Field Gel Electrophoresis*) mostrando variação cariotípica observada em derivados de JAY270 (FDY1-FDY11).** Os derivados apresentam polimorfismos de tamanho nos cromossomos em relação ao parental JAY270, devido à recombinação mitótica. Os números localizados à esquerda correspondem aos dezesseis cromossomos de JAY270 (Duarte 2010).

## MATERIAIS E MÉTODOS

### Meios de cultura e linhagens de levedura

Para ensaios fermentativos de competição entre linhagens industriais de levedura, foi utilizado xarope diluído como meio de cultura, contendo 18% de xarope de cana-de-açúcar, fornecido pela Usina Granelli, localizada em Charqueada, SP. Ao final dos ciclos fermentativos, as culturas foram plaqueadas em YPD, meio de cultura padrão utilizado

para o crescimento das linhagens de *S. cerevisiae*, contendo 1% de extrato de levedura, 2% de peptona, 2% de glicose e 2% ágar. O meio YPD foi complementado com o antibiótico geneticina (G418) a uma concentração de 200 µg/mL ou higromicina em concentração de 300 µg/mL, após autoclavagem. As linhagens de levedura utilizadas nesse estudo estão descritas na Tabela 1.

**Tabela 1** - Linhagens de levedura *Saccharomyces cerevisiae* utilizadas nesse estudo

Descrições:		Referências:
<b>Diploides:</b>		
JAY270:	Isolado selvagem da linhagem industrial PE-2, purificado de estoque comercial	Argueso <i>et al.</i> 2009
FDY1 a FDY11:	Onze isolados que apresentam variação cariotípica em relação ao parental JAY270, coletados ao final de experimento de fermentação industrial de 60 dias	Duarte 2010
FDY1H a FDY11H:	Derivadas das linhagens FDY1 a FDY11, portando o gene de resistência para o antibiótico higromicina ( <i>HphMX4</i> ), inserido a 500 pb do centrômero do Cromossomo V (CEN5)	Nossa coleção
FDY6K:	Derivada da linhagem FDY6, portando o gene de resistência para o antibiótico geneticina ( <i>KanMX4</i> ), inserido a 500 pb do centrômero do Cromossomo V (CEN5)	Nossa coleção
JAY270H:	Derivada da linhagem JAY270 portando o gene de resistência para o antibiótico higromicina ( <i>HphMX4</i> ), inserido a 500 pb do centrômero do Cromossomo V (CEN5)	Nossa coleção
JAY270K:	Derivada da linhagem JAY270, portando o gene de resistência para o antibiótico geneticina ( <i>KanMX4</i> ), inserido a 500 pb do centrômero do Cromossomo V (CEN5)	Nossa coleção

## Ensaio fermentativo de competição entre linhagens LOH e parental

Os onze derivados da linhagem JAY270 portadores de variação cariotípica (FDY1-FDY11), descritos anteriormente (Figura 1), foram comparados e analisados

fenotipicamente através de ensaios fermentativos de competição direta com a linhagem parental JAY270.

Para que a realização desse experimento fosse possível, primeiramente a linhagem parental JAY270 e os derivados LOH foram marcados com resistência à drogas diferentes, permitindo assim monitorar a razão relativa celular LOH para parental na população, durante a co-cultura. Desse modo, foi inserido um gene que confere resistência ao antibiótico higromicina no genoma de FDY1-FDY11 e um gene que confere resistência à geneticina no genoma de JAY270. As linhagens LOH portadoras do marcador de resistência à higromicina foram chamadas FDY1H a FDY11H. Já a linhagem JAY270 portadora do gene de resistência ao antibiótico geneticina foi chamada de JAY270K.

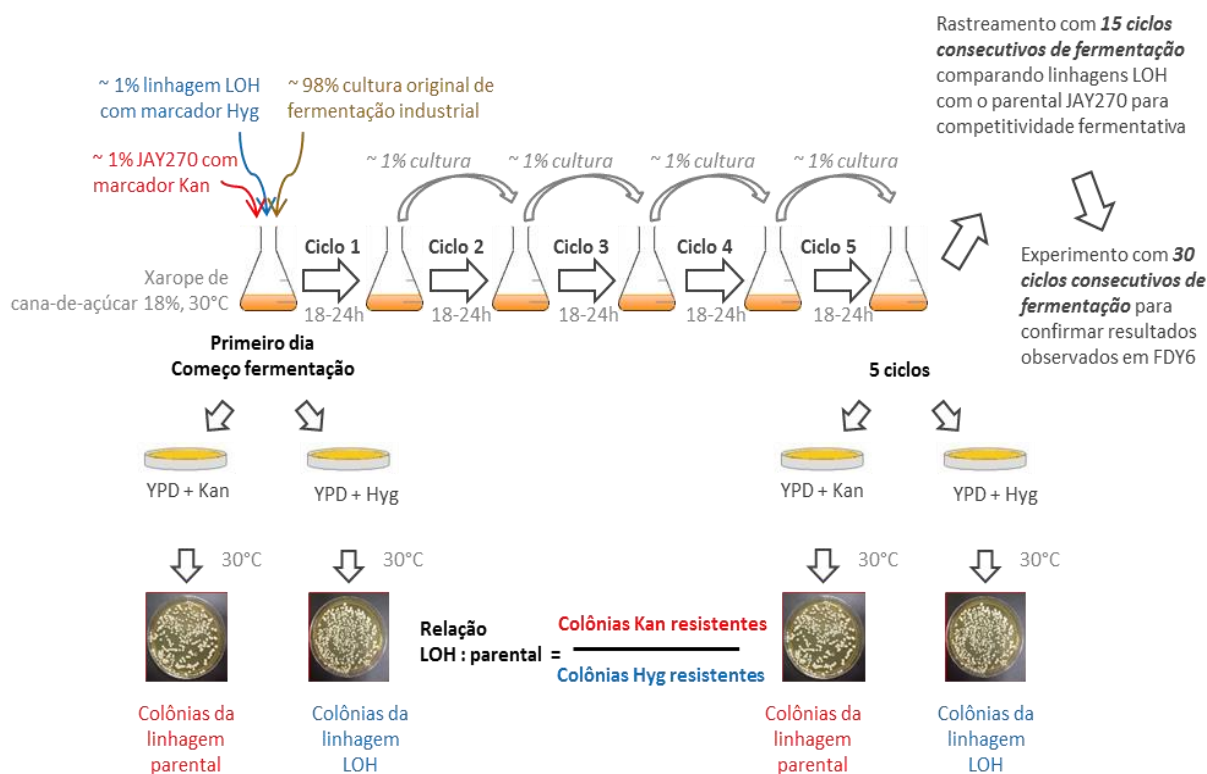
Anteriormente ao preparo do inóculo inicial, cada linhagem FDYH, JAY270K e JAY270H foram crescidas isoladamente em 5ml de meio YPD sem antibiótico, a temperatura de 30°C *overnight*.

O inóculo inicial foi preparado através da mistura de mesma porcentagem de JAY270 pura (JAY270K) com cada um dos onze derivados (FDY1H-FDY11H) e também com uma cultura original de fermentação industrial (razão relativa celular LOH:parental inicial de 1:1). Sendo assim, 1% de cada cultura das linhagens FDY1H-11H foi misturada com 1% de cultura JAY270K e 98% da cultura original de dornas de fermentação. Dessa mistura total, 1% foi inoculado em 50 ml de xarope de cana-de-açúcar diluído e autoclavado. Também realizamos uma competição controle entre JAY270K e JAY270H, com inóculo preparado da mesma maneira.



A cultura original de dornas de fermentação utilizada contém leveduras selvagens, em sua maioria PE-2, além de outros microorganismos não identificados e foi coletada em um experimento de fermentação em escala industrial realizado em uma destilaria piloto da ESALQ-USP durante doutorado de um aluno do grupo (Carvalho-Netto 2012).

Ao final de cada ciclo fermentativo (~24 horas), uma pequena amostra das culturas (500 µl) era transferida para meio fresco, iniciando um novo ciclo. A cada cinco ciclos, amostras foram plaqueadas em meio YPD contendo os antibióticos específicos e a razão LOH:parental foi determinada através de contagem das colônias após crescimento a 30°C durante 3 dias. Primeiramente, a competitividade das onze linhagens foi avaliado durante fermentação contínua de xarope de cana-de-açúcar em quinze ciclos. Os resultados foram então confirmados em um experimento com trinta ciclos fermentativos. Os ensaios foram realizados em uma estufa a 30°C com agitação de 150 rpm. Uma esquematização dos experimentos é apresentada na Figura 2.



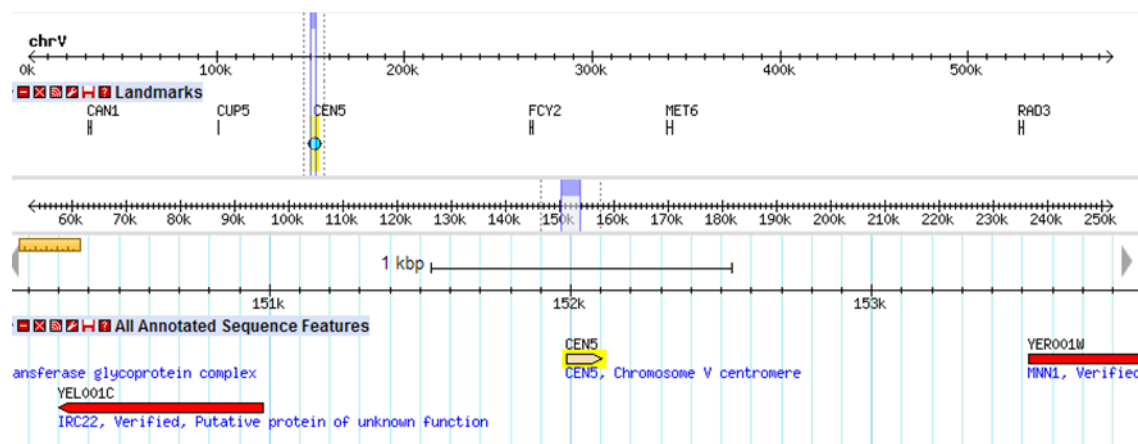
**Figura 2 – Experimentos de competição de longo prazo para testar consequências fenotípicas de perda de heterozigidade (LOH) em linhagens de PE-2.** Representação esquemática dos ensaios fermentativos de competição realizados entre linhagens LOH e a linhagem parental JAY270.

## RESULTADOS

Para investigar se os rearranjos cromossômicos que ocorrem no genoma da linhagem industrial PE-2 durante a produção de etanol estão relacionados com alguma alteração fenotípica, foram analisados onze derivados de JAY270 portadores de variação estrutural cromossômica com perda de heterozigidade (FDY1-FDY11), identificados e selecionados previamente pelo nosso grupo de pesquisa (Duarte 2010).

Esses derivados foram comparados com a linhagem parental JAY270 através de ensaios de competição. Para isso, a linhagem JAY270 e seus derivados foram marcados

com genes de resistência a antibióticos específicos, conforme descrito em “materiais e métodos”. Os genes de resistência, *KanMX4* e *HphMX4*, foram inseridos em cópia única, a 500 pb do centrômero do Cromossomo V (*CEN5*), entre o gene *IRC22* e o *CEN5* das linhagens, em uma região que não contém genes (Figura 3). Nesta localização, a probabilidade de perda dos marcadores por recombinação homóloga é praticamente nula.

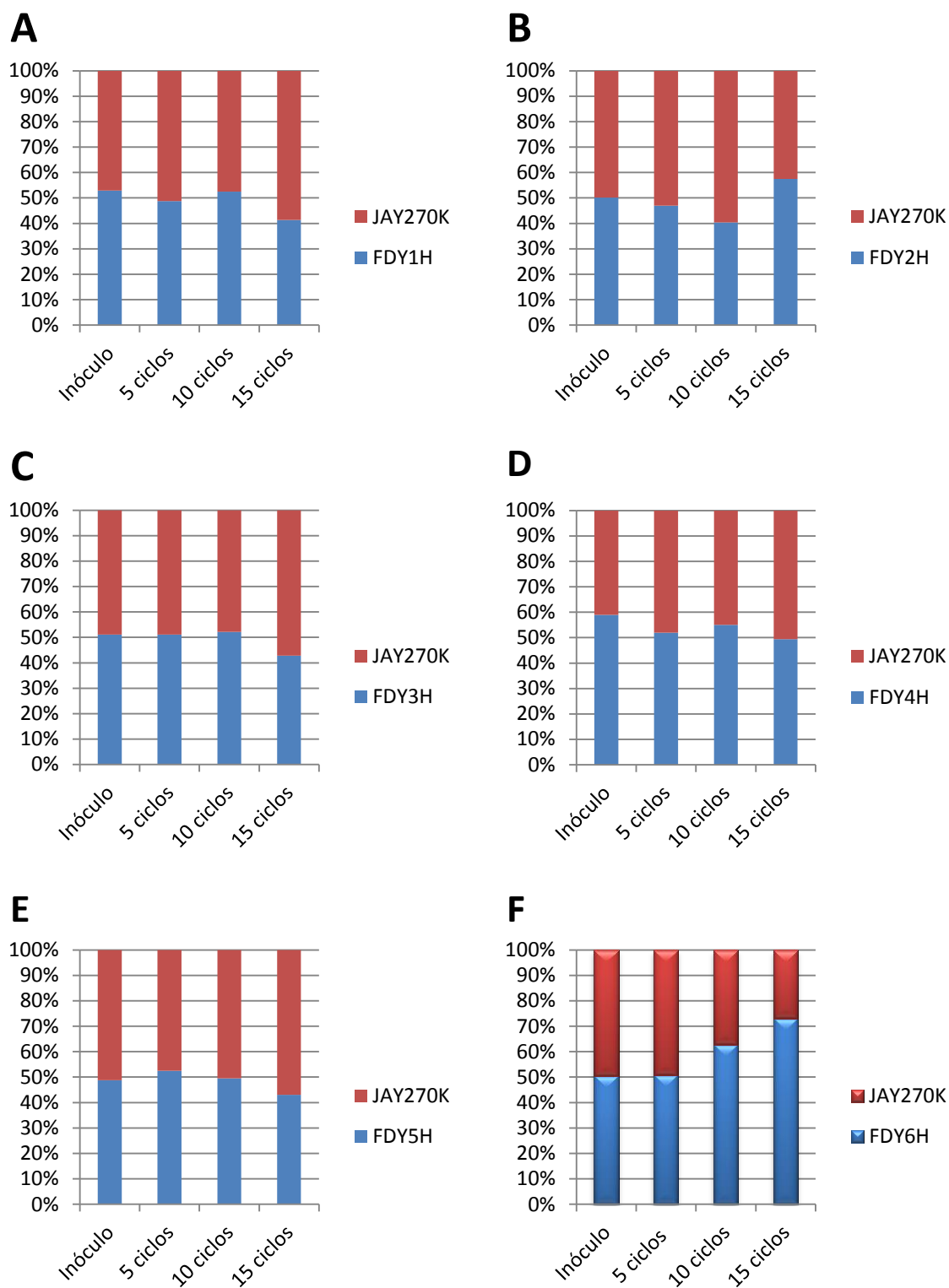


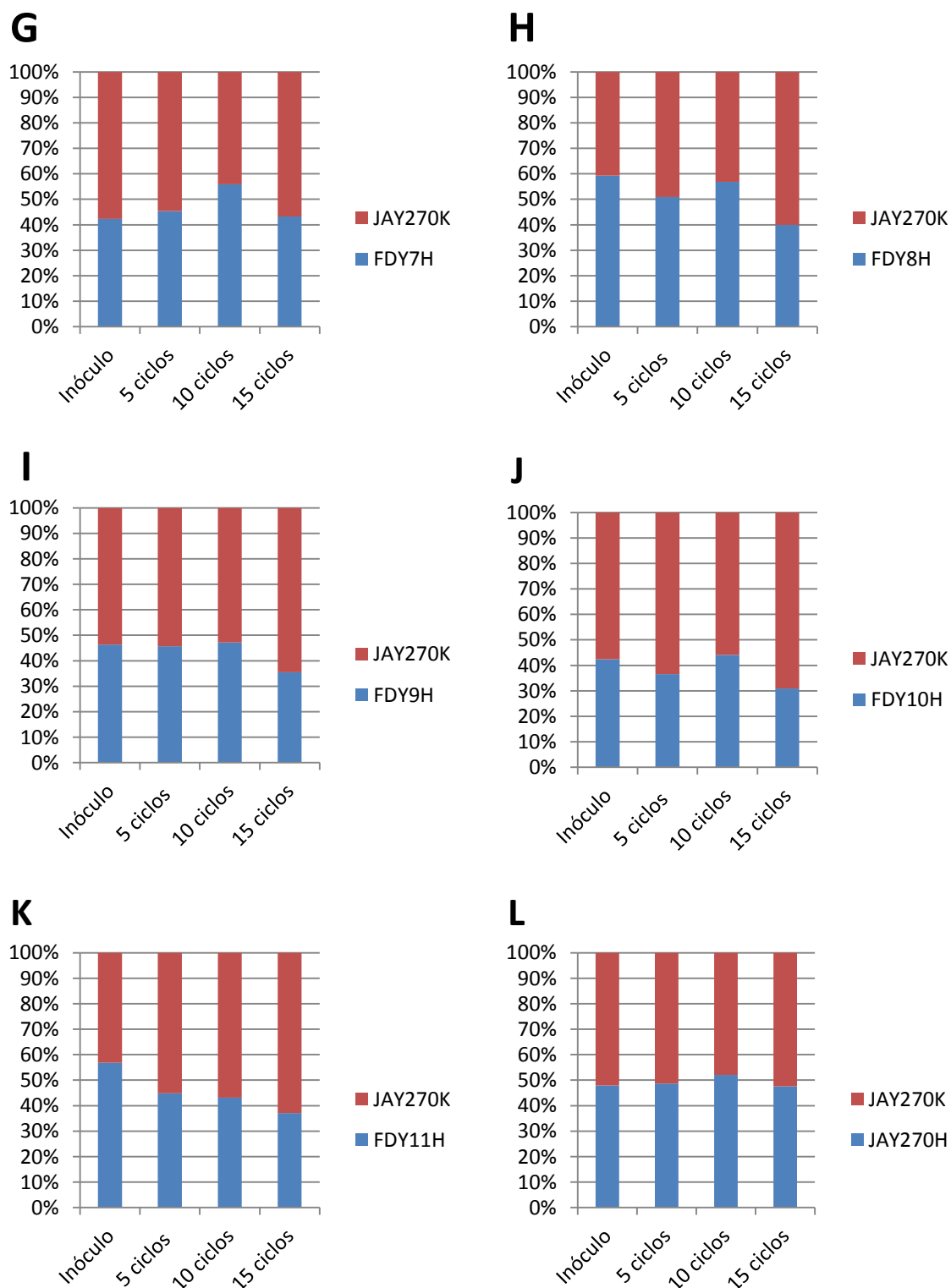
**Figura 3 - Local de inserção das marcas de seleção no genoma das linhagens LOH e JAY270.** Os genes de resistência à higromicina (*HphMX4*) ou geneticina (*KanMX4*) foram inseridos nos genomas das linhagens LOH e JAY270 respectivamente. As marcas de seleção foram inseridas a 500 pb do centrômero 5 (*CEN5*) das linhagens, entre o gene *IRC22* e *CEN5*.

Como descrito em “materiais e métodos” primeiramente realizamos um rastreamento, competindo cada uma das onze linhagens LOH (FDY1H-FDY11H) com a linhagem parental (JAY270K) em ensaios fermentativos contendo quinze ciclos consecutivos. A razão celular LOH:parental foi monitorada a cada cinco ciclos fermentativos.

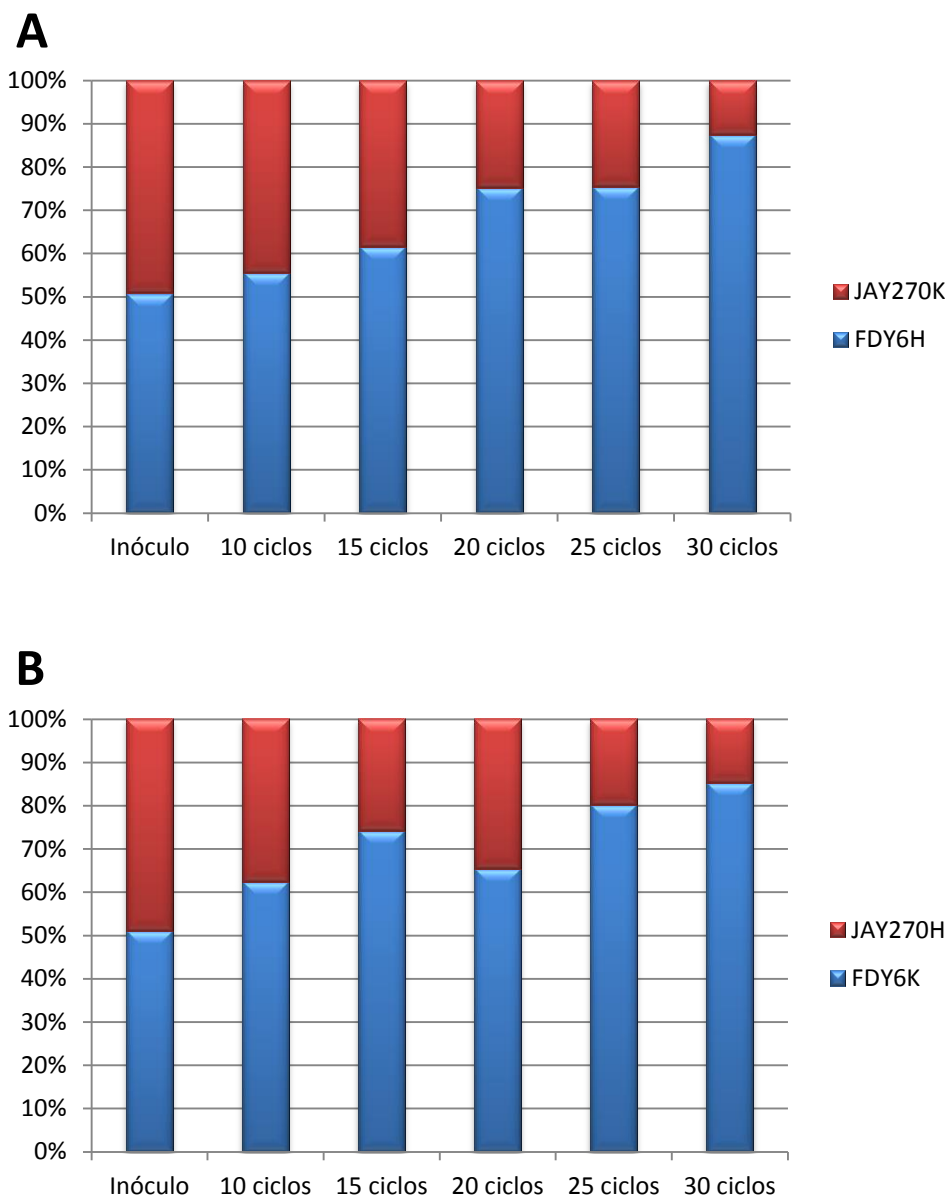
A razão relativa celular LOH:parental inicial de 1:1 se manteve aproximadamente constante até o final dos quinze ciclos fermentativos para dez das onze linhagens LOH avaliadas (Figura 4). Dentre todos os ensaios realizados, somente uma linhagem (FDY6), apresentou uma dominância na fermentação com relação à linhagem parental JAY270

(Figura 4F). O experimento controle de competição entre JAY270K e JAY270H demonstrou a não interferência dos marcadores de resistência no estudo (Figura 4L).





**Figura 4 - Competição direta: linhagens LOH X parental JAY270 (rastreamento com quinze ciclos).** Experimentos de competição com quinze ciclos fermentativos consecutivos, comparando-se linhagens LOH com o parental JAY270.



**Figura 5 – Linhagem FDY6 (LOH) X Linhagem Parental JAY270.** Ensaios de competição contendo 30 ciclos consecutivos de fermentação para confirmar resultados obtidos para FDY6.

Para confirmar os resultados obtidos para FDY6, repetimos o mesmo ensaio de competição realizado no rastreamento, mas dessa vez com trinta ciclos consecutivos de fermentação, para essa linhagem. Para eliminar a interferência de plaqueamento com os antibióticos utilizados, realizamos dois ensaios de competição, um entre a linhagem FDY6

marcador de resistência à higromicina (FDY6H) e linhagem parental marcada com gene de resistência à geneticina (JAY270K) e outro ensaio entre FDY6 com marcador de resistência à geneticina (FDY6K) e linhagem parental com marcador de resistência à higromicina (JAY270H).

A fermentação contínua de trinta ciclos confirmou o primeiro resultado obtido: linhagem LOH FDY6 apresentou uma dominância significativa em relação à linhagem parental JAY270 (Figura 5).

## DISCUSSÃO

Nesse capítulo, testamos a competitividade de linhagens PE-2 com variações cariotípicas (FDY1-11), previamente estudadas pelo nosso grupo de pesquisa (Duarte 2010), para avaliar se LOH pode ser associado com alterações fenotípicas. As onze linhagens LOH foram avaliadas fenotipicamente durante fermentação contínua de xarope de cana-de-açúcar em competição direta contra a linhagem parental JAY270, que teve seu genoma molecularmente estudado e demonstrou ser altamente heterozigota (Argueso *et al.* 2009).

Sabe-se que o processo de fermentação industrial para produção de etanol pode ser estressante para leveduras presentes nas dornas, com sucessões intensas de linhagens e alta competição. A estabilidade das populações depende da utilização de linhagens resistentes e bem adaptadas ao processo industrial e é importante para ajudar a manter a taxa de produção de etanol alta e estável (Basso *et al.* 2008; Amorim *et al.* 2011). As linhagens FDY1-11 foram coletadas no final de experimento industrial contendo 60 ciclos fermentativos (Duarte 2010), demonstrando que são linhagens competitivas e resistentes,

conseguindo permanecer até o final desse processo. Acreditamos então que, nesse experimento de competição direta com o parental, as variações estruturais cromossômicas com perda de heterozigosidade apresentadas em FDY1-11 se demonstrariam neutras ou benéficas às linhagens, pois variações negativas poderiam eliminar linhagens no processo.

Conforme esperado, as variações cariotípicas se demonstraram neutras para a competitividade da maioria das linhagens LOH quando comparadas ao parental, sendo benéfica à uma delas. Dentre as onze linhagens avaliadas (FDY1-11), somente uma (FDY6) apresentou significativa dominância durante a fermentação relativamente ao seu parental sendo, portanto, que a perda de heterozigosidade que ocorre no genoma de JAY270 durante a fermentação pode ser associada ao ganho de competitividade apresentado em FDY6 nos ensaios de competição.

## CONCLUSÕES

- ✓ A variação estrutural cromossômica em JAY270 pode ser associada com efeitos benéficos à competitividade da linhagem durante a fermentação.
- ✓ Outros estudos estão sendo realizados pelo grupo para melhor compreensão e caracterização genética da dominância apresentada em FDY6 e também para localizar regiões genômicas envolvidas.



## REFERÊNCIAS

- Amorim, H. V, M. L. Lopes, J. V. de Castro Oliveira, M. S. Buckeridge, and G. H. Goldman, 2011 Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 1267–1275.
- Argueso, J. L., M. F. Carazzolle, P. A. Mieczkowski, F. M. Duarte, O. V. C. Netto *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19: 2258–70.
- Basso, L. C., H. V de Amorim, A. J. de Oliveira, and M. L. Lopes, 2008 Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res.* 8: 1155–63.
- Carvalho-Netto, O. V., 2012 Fisiologia da fermentação alcoólica: análise da expressão gênica de uma linhagem industrial de *Saccharomyces cerevisiae* durante o processo fermentativo: Universidade Estadual de Campinas. Tese de Doutorado. 165p.
- Duarte, F. M., 2010 Análise da variação cariotípica de recombinação em leveduras industriais (*Saccharomyces cerevisiae*) durante o processo de fermentação alcoólica: Universidade Estadual de Campinas. Tese de Mestrado. 68p.



## **Conclusões Gerais**



## Conclusões Gerais

Durante o desenvolvimento desses estudos, realizamos diferentes experimentos envolvendo as linhagens industriais *Saccharomyces cerevisiae* diploide PE-2/JAY270 e haploide JAY291 (derivado de JAY270), dos quais temos as seguintes conclusões:

Através de mapeamento genômico por novos genes de floculação, analisando segregantes F2 de um cruzamento entre a linhagem industrial JAY291 e de laboratório S288c, observamos que esse fenômeno é controlado principalmente por três genes: *FLO1*, *FLO8* e *MSS11*, necessários para a ocorrência de floculação.

Além desses três genes principais, encontramos mais doze loci associados com floculação, que contém genes de efeito menor, mas também atuantes sobre a característica. Observamos ainda que muitos loci associados estavam localizados em regiões subteloméricas.

Existem genes identificados nos loci, relacionados com outros fenótipos de adesão celular em estudos descritos na literatura como biofilme, pseudohifas e crescimento invasivo, que foram considerados no presente trabalho como candidatos para associação com floculação.

Em outros estudos, vimos que a transição da típica morfologia de colônias lisas de JAY270 para rugosas, acompanhada de alta sedimentação, é resultado de um defeito de separação celular. O gene *ACE2* foi identificado como responsável por essa transição.

Nossas descobertas mostraram que a linhagem JAY270 é heterozigota para uma mutação em *frameshift* nesse gene (*ACE2/ace2-7A*), sendo que o fenótipo rugoso ocorre por recombinação mitótica associada com perda de heterozigosidade (LOH), produzindo células com duas cópias mutadas desse gene (*ace2-7A /ace2-7A*).

O defeito detectado *ace2-7A* foi reparado no genoma da JAY270. A linhagem desenvolvida PE-2/ARY103 passou a ter duas cópias do alelo selvagem de *ACE2* (*ACE2/ACE2*) e portanto incapaz de se tornar rugosa por recombinação e consequente perda de heterozigosidade.

Em outros experimentos, avaliamos as consequências fenotípicas da perda de heterozigosidade através de competição de linhagens LOH (derivadas de JAY270) em ensaios de competição direta com o parental. A linhagem FDY6 apresentou dominância fermentativa sobre JAY270.

Os estudos realizados colaboram para um melhor entendimento da genômica funcional da linhagem PE-2/JAY270.

## **Anexo I**

### **A simple and effective set of PCR-based molecular markers for the monitoring of the *Saccharomyces cerevisiae* cell population during bioethanol fermentation**

Esse anexo contém um manuscrito aceito para publicação na revista *Journal of Biotechnology*, no qual tive participação como co-autora pela realização de alguns experimentos de genotipagem de linhagens por PCR.







# A simple and effective set of PCR-based molecular markers for the monitoring of the *Saccharomyces cerevisiae* cell population during bioethanol fermentation

Osmar V. Carvalho-Netto<sup>a,\*</sup>, Marcelo F. Carazzolle<sup>a</sup>, Aline Rodrigues<sup>a</sup>, Welbe O. Bragança<sup>a</sup>, Gustavo G.L. Costa<sup>a</sup>, Juan Lucas Argüeso<sup>b</sup>, Gonçalo A.G. Pereira<sup>a</sup>

<sup>a</sup> Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas-SP, Brazil

<sup>b</sup> Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, USA

## ARTICLE INFO

**Article history:**  
Received 26 May 2013  
Received in revised form 19 August 2013  
Accepted 21 August 2013  
Available online xxx

**Keywords:**  
*Saccharomyces cerevisiae*  
PCR  
Genotyping  
Bioethanol

## ABSTRACT

One of the defining features of the fermentation process used in the production of bioethanol from sugarcane feedstock is the dynamic nature of the yeast population. Minisatellite molecular markers are particularly useful for monitoring yeast communities because they produce polymorphic PCR products that typically display wide size variations. We compared the coding sequences derived from the genome of the sugarcane bioethanol strain JAY270/PE-2 to those of the reference *Saccharomyces cerevisiae* laboratory strain S288c, and searched for genes containing insertion or deletion polymorphisms larger than 24 bp. We then designed oligonucleotide primers flanking nine of these sites, and used them to amplify differentially sized PCR products. We analyzed the banding patterns in the most widely adopted sugarcane bioethanol strains and in several indigenous yeast contaminants, and found that our marker set had very good discriminatory power. Subsequently, these markers were used to successfully monitor the yeast cell populations in six sugarcane bioethanol distilleries. Additionally, we showed that most of the markers described here are also polymorphic among strains unrelated to bioethanol production, suggesting that they may be applied universally in *S. cerevisiae*. Because the relatively large polymorphisms are detectable in conventional agarose gels, our method is well suited to modestly equipped on-site laboratories at bioethanol distilleries, therefore providing both cost and time savings.

© 2013 Published by Elsevier B.V.

## 1. Introduction

In Brazil and other tropical countries that grow sugarcane, ethanol biofuel (bioethanol) is produced through fermentation of sucrose by the yeast *Saccharomyces cerevisiae*, primarily using a fed-batch process (Wheals et al., 1999). This high moisture/high sugar feedstock brings into the industrial plant a substantial microbial load of wild yeasts and bacteria (Amorim et al., 2011). A heat treatment is applied to reduce the number of viable contaminants, but inevitably, some find their way into the fermentation tanks where they have the opportunity to flourish. After the fermentation of each batch of sugarcane juice is completed, the cell population is recycled to ferment the subsequent batches, and this practice is repeated continuously for the duration of the harvest season

(8–9 months). The effects of the steady influx of microbial contaminants are exacerbated by the cell recycling regimen widely adopted by the industry to achieve a high rate of feedstock processing: i.e., a dense yeast inoculum (~10% cells by volume) leads to faster fermentations (~10 h). Consequently, the microbial community in most distilleries is heterogeneous and highly dynamic. The starter strains inoculated in the beginning of the season (habitually baker's yeast due to their wide market availability and low cost) are rapidly out-competed and replaced by more robust indigenous strains that can endure the stressful conditions encountered during the fermentation (da Silva et al., 2005).

Over the past fifteen years, indigenous strains that combine excellent fermentative characteristics and high adaptation to the industrial environment have been selected from Brazilian bioethanol distilleries and are used today as starter strains. Among those, the PE-2 and CAT-1 strains are the most widely adopted (Basso et al., 2008). However, PE-2 and CAT-1 are exceptions among their indigenous counterparts, which in their majority display undesirable fermentation characteristics, and often result in significant productivity losses if they are allowed to establish a foothold in a distillery (Basso et al., 2008; Wheals et al., 1999).

\* Corresponding author at: Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas-SP 13083-970, Brazil. Tel.: +55 19 3521 6237.  
E-mail addresses: [osmar@ige.ibi.unicamp.br](mailto:osmar@ige.ibi.unicamp.br), [osmar Netto@hotmail.com](mailto:osmar Netto@hotmail.com) (O.V. Carvalho-Netto).

Therefore, effective methods are needed to monitor the microbial population to identify the presence of invaders in a timely fashion.

A wide array of methodologies has been previously described for the molecular identification of *S. cerevisiae* strains including ribosomal DNA or ITS sequencing (Kurtzman and Robnett, 1998), RAPD (Bovo et al., 2009; Gomes et al., 2000), mtDNA RFLP (Querol et al., 1992), PCR-RFLP (Schuller et al., 2004, 2007), real-time PCR (Hierro et al., 2006; Martorell et al., 2005), microsatellite and interdelta sequence amplification (Legras and Karst, 2003; Legras et al., 2005; Schuller et al., 2007; Vaudano and Garcia-Moruno, 2008), minisatellite intragenic tandem repeats (Marinangeli et al., 2004; Verstrepen et al., 2005), and karyotyping by pulsed-field gel electrophoresis (PFGE) (Basso et al., 2008; Schuller et al., 2004, 2007). All of these methods can reveal polymorphisms between industrial yeast strains and the indigenous contaminants. However, a limitation of several of these approaches is that they sometimes require relatively sophisticated equipment and expertise that are rarely available at bioethanol distilleries. For example, the analysis of microsatellite markers occasionally requires high-resolution capillary gel electrophoresis to resolve small polymorphisms; methods that involve restriction fragmentation or RAPD often generate complex patterns with multiple bands that can be difficult to analyze; and karyotyping requires a dedicated pulse field apparatus. In part due to their technical inability to carry out some of these procedures, most bioethanol distilleries opt to outsource the molecular monitoring of the yeast population, even though they often have on-site quality control laboratories that are reasonably capable.

The standard monitoring assay used in Brazil since the mid 1990s has been karyotyping by PFGE, and microsatellite markers have increased in use in recent years. While these have proven useful, they are normally conducted at third party laboratories, and several days may elapse between sample collection, shipment, colony isolation, analysis, and the final read-out. By that point in time, the composition of the yeast population in the distillery may already have changed significantly.

One of the most straightforward methods cited above is the detection of relative large genetic polymorphisms at minisatellite loci, whose analysis does not require specialized electrophoresis equipment or the expertise to operate them. In addition, the banding patterns tend to be simpler (one or two bands per sample), while still providing substantial discriminatory power between strains (Marinangeli et al., 2004; Verstrepen et al., 2005). As stated above, simplicity is a key factor to the dissemination of new analytical technologies at industrial laboratories, particularly in remote sugarcane production regions. Therefore, this class of molecular markers represents a viable alternative to PFGE and microsatellites because it would allow the yeast population monitoring to be carried out on-site, and have the added advantage of a shorter time interval between sampling and the final result.

We have previously described the genome sequence of a haploid derivative (JAY291) from the JAY270/PE-2 bioethanol strain (Argueso et al., 2009). We took advantage of our sequence data to design PCR markers at loci that show relatively large size polymorphisms in bioethanol strains, primarily at minisatellite loci. We carried out a genome-wide search for length polymorphic regions within genes, by comparing the JAY291 haplotype sequence to those of S288c and other sequenced *S. cerevisiae* strains. This approach allowed us to develop markers that can positively identify JAY270/PE-2 using a small number of loci. We found that these loci are also polymorphic in other major bioethanol strains, as well as in uncharacterized indigenous yeast contaminants. The markers presented here should be immediately useful to the large number of distilleries that use strain JAY270/PE-2, as well as other *S. cerevisiae* strains.

## 2. Materials and methods

### 2.1. Strains and culture

The *S. cerevisiae* strains used in this study were predominantly derived from Brazilian bioethanol distilleries or *cachaça* still (listed in Supplementary Table 1; (Argueso et al., 2009; Basso et al., 2008; Campos et al., 2010; Gomes et al., 2000; Kane and Roth, 1974; Mortimer et al., 1994; Novo et al., 2009; Wei et al., 2007; Winston et al., 1995)). The main industrial strains used in sugarcane bioethanol production are of indigenous origin, and were all isolated as high productivity invaders. In addition, fermentation samples were collected from six different Brazilian bioethanol distilleries during the 2011 sugarcane harvest season to monitor the yeast population. Samples collected at bioethanol production plants were streaked onto YPD solid medium [yeast extract 1% (w/v), peptone 2% (w/v), glucose 2% (w/v) and agar 2% (w/v)] for single colony isolation, and incubated at 30 °C for 2 days. Colonies were then cultured overnight in 4 mL YPD liquid medium at 30 °C for genomic DNA extraction.

### 2.2. Identification and selection of length polymorphic regions within genes

In order to identify polymorphic regions within genes, the coding regions from the *S. cerevisiae* reference genome (strain S288c) were aligned to the assembled genome sequence of JAY291 using Exonerate (Slater and Birney, 2005), applying stringent parameters to produce accurate sequence alignments (-best 1 -model est2genome -percent 20 -forcecgtag TRUE -quality 80 -refine region). A custom PERL script was created to parse the Exonerate output, identifying the insertion and deletion polymorphisms present in the output. Only genes containing regions larger than 24 consecutive gaps and with gap lengths divisible by three (unaltered reading frames) were selected, resulting in an initial list of 27 candidate genes with usable length polymorphic regions. The sequences of these genes were then analyzed in the genomes of the unrelated *S. cerevisiae* strains YJM789 (Wei et al., 2007), AWRI1631 (Borneman et al., 2008), and RM11-1a (*S. cerevisiae* RM11-1a Sequencing Project; <http://www.broad.mit.edu/>) through nucleotide similarity searches in GenBank using BLASTn. Of the 27 initial sequences analyzed, the nine genes (Table 1) that displayed the highest degree of length variability between strains were selected as the most suitable for use as PCR marker loci.

It is important to note that genes and other sequences containing short internal repeats often coincided with the ends of the contigs obtained during the JAY291 genome assembly (Argueso et al., 2009), as is typical for assemblies derived from short next generation sequencing reads. Only fully assembled segments of the JAY291 genome were used in the search for length polymorphisms, therefore it is likely that additional polymorphic regions exist, but were not identified in the present study (see Section 3).

### 2.3. Amplification of length polymorphic regions

Primer3 software (Rozen and Skaletsky, 2000) was used to design optimal oligonucleotide primers to amplify the selected polymorphic regions. In seven of the nine loci the primers were within the coding sequence. The exceptions were YKL207C in which the reverse primer was downstream of the coding sequence, and YDR504C with both primers located outside the coding region. The primer sequences used to amplify the target regions are shown in Table 1.

Yeast genomic DNA was extracted using a phenol-chloroform protocol (Ausubel et al., 1998). PCR was carried out in 10 µL volume reactions containing 20 ng of genomic DNA, 0.25 mM dNTP (each),



**Table 1**

List of genes identified in this study, with PCR product sizes and primers flanking the length polymorphic regions.

ORF identifier	Gene name	PCR product size (bp)		5'–3' primer sequences			Melting temperature (°C)
		S288c	JAY291	$\Delta^d$	Forward	Reverse	
YBL046W	PSY4	514	541	27	GCAGTTGGCAGGTTTGTAT	GCCTGCCTATCTTCGTCTATC	55
YDR299W	BFR2	433	460	27	AGCGGATCAAAATTCGATA	TGCATGCTTTTCTGTTGAG	52
YDR504C	SPG3	564	600	36	TCGACAACTTTCAGATTGCTG	AATGGAAGCGGCTGTAGAA	55
YDR505C	PSP1	425	497	72	ACAGTTTCGCCGGAATACAC	CCATGTTCTTCGCGTCATAA	52
YFL024C	EPL1	385	415	30	ACGATTCCAAATACGACGAA	TTCTGTTTCGCTTCTGAATTG	57
YKL163W	PIR3 <sup>a</sup>	391	553 <sup>a</sup>	162	TGTCCGCTCATCTAAAGCAA	TGTAATTTGGGATCGACGAC	54
YKL201C	MNN4 <sup>c</sup>	699	631 <sup>b</sup>	68	TAGACCTTTTTCGCCCACT	ATTACACGATTCGCTCGAA	57
YLL021W	SPA2 <sup>c</sup>	655	817	162	GAAATGACGATGACAGACGA	AGGACTCGCTTTCCTTACC	55
YPR143W	RRP15	562	601	39	GGGCTCCAAGCACAGAGTAG	CTTCTGAGCGCTGCTCTCT	57

<sup>a</sup> Re-sequenced PCR product; The predicted PCR product size for YKL163W according to the first JAY291 genomic assembly was 337 bp.<sup>b</sup> Re-sequenced PCR product; The predicted PCR product size for YKL201C according to the first JAY291 genomic assembly was 582 bp.<sup>c</sup> Three polymorphic loci also identified in the Verstrep et al. (2005) gene set [17].<sup>d</sup> Absolute difference in size between the S288c and JAY291 PCR products.

181 1× PCR buffer, 3.5 mM MgCl<sub>2</sub>, 0.05 U/μL of Taq DNA polymerase,  
 182 1.5 pmol/μL of each primer. The amplification program consisted  
 183 of 5 min denaturation at 94 °C, followed by 40 cycles of 45 s at 94 °C,  
 184 40 s at primer-specific annealing temperature (Table 1), and 45 s at  
 185 72 °C. A 5 min final extension was performed at 72 °C. PCR ampli-  
 186 cons were resolved in conventional electrophoresis on 2% agarose  
 187 gels in 1× TAE buffer (40 mM Tris–acetate, 2 mM Na<sub>2</sub>EDTA·2H<sub>2</sub>O, pH  
 188 8.5). Gels were stained with ethidium bromide, visualized under UV  
 189 light, and photographed. The size and number of the detected bands  
 190 were scored relative to a molecular weight ladder. We have also  
 191 successfully amplified the marker PCR products from crude yeast  
 192 DNA template preparations and using shorter cycling programs (*i.e.*  
 193 colony PCR; data not shown), therefore simpler and faster custom  
 194 DNA preparation and PCR protocols can be employed.

#### 195 2.4. *In silico* PCR analysis

196 The S288c sequences of the nine selected loci, in their respective  
 197 regions delimited by the primer pairs listed in Table 1 were used  
 198 as reference for *in silico* PCR analysis. The comparison of sequence  
 199 lengths at each locus in multiple *S. cerevisiae* strains was con-  
 200 ducted by BLASTn searches using the complete genomic sequences  
 201 deposited at the Sanger Institute, SGD and NCBI databases. Exper-  
 202 imental validation of the *in silico* analysis was conducted using the  
 203 PCR protocol described above.

#### 204 2.5. Evaluation of discriminatory power

205 The discriminatory power of each PCR marker locus was mea-  
 206 sured by the Polymorphic Index Content (PIC) parameter, using  
 207 the algorithm:  $PIC = 1 - \sum (p_i^2)$ , where  $p_i$  is the frequency of the  
 208  $i^{th}$  allele detected (Smith et al., 1997). A marker with PIC = 0 is  
 209 monomorphic, while a marker with PIC = 1 is very highly discrimi-  
 210 native.

#### 211 2.6. Re-sequencing of specific loci in JAY291

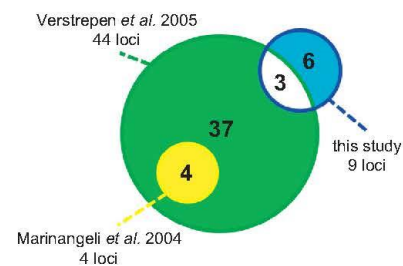
212 As part of this work, we re-sequenced the YKL163W and YKL201C  
 213 loci from haploid strain JAY291 because they were found to be  
 214 mis-assembled in the original JAY291 genome report. The new  
 215 correct sequences were used to edit the existing JAY291 contigs  
 216 deposited in GenBank under accession numbers ACFL01000239.2  
 217 and ACFL01000006.2, respectively. The YFL024C, YKL201C, and  
 218 YLL021W loci from the sibling haploid JAY292 were deposited under  
 219 GenBank accession numbers JX101677, JX101678, and JX101679,  
 220 respectively. All of these sequences were obtained by conventional  
 221 Sanger sequencing of PCR products generated with the primers in  
 222 Table 1.

### 223 3. Results and discussion

#### 224 3.1. Identification of length polymorphic regions within genes

225 The goal of this study was to develop simple, yet effective, PCR  
 226 markers for the molecular identification of Brazilian bioethanol  
 227 yeast strains. We compared coding sequences from the genome  
 228 of the standard laboratory strain S288c (Goffeau et al., 1996)  
 229 to those assembled from the genome sequencing of strain JAY291, a  
 230 haploid derivative of JAY270/PE-2 (Argueso et al., 2009). We then  
 231 refined the search for length polymorphic loci by examining candi-  
 232 date regions in three other unrelated sequenced strains (YJM789,  
 233 AWRI1631, and RM11-1a; Section 2). Nine genes were identified  
 234 that contained sufficiently large continuous length polymorphic  
 235 regions that could be discriminated by conventional agarose gel  
 236 electrophoresis, therefore simplifying the readout of the analysis.  
 237 Interestingly, the list of loci identified through our targeted S288c  
 238 vs. JAY291 comparison approach (Table 1) had relatively little over-  
 239 lap with similar loci described earlier (Fig. 1). Three of the genes  
 240 identified in our study (YKL163W, YKL201C, and YLL021W) had been  
 241 found previously in a systematic survey of the S288c reference  
 242 genome for genes containing tandem repeats within their coding  
 243 sequences (Verstrep et al., 2005), and none were found in com-  
 244 mon with another study that focused on potentially polymorphic  
 245 cell wall genes (Marinangeli et al., 2004). Conversely, our approach  
 246 identified six genes that were not found in these earlier studies.

247 To better understand the reasons behind these differences we  
 248 systematically repeated the sequence alignments between S288c



**Fig. 1.** Relationships between the polymorphic loci described previously and those identified in this study.

Venn diagram showing the number of loci with internal size polymorphisms found in this study (blue circle) and those previously described by Marinangeli et al. (2004) (yellow) and Verstrep et al. (2005) (green). The three genes identified in both this study and (Verstrep et al., 2005) (white) are YKL163W (PIR3); YKL201C (MNN4); and YLL021W (SPA2). The area of the circles is roughly proportional to the number of genes. The full list of genes is presented in Supplementary Table 2.

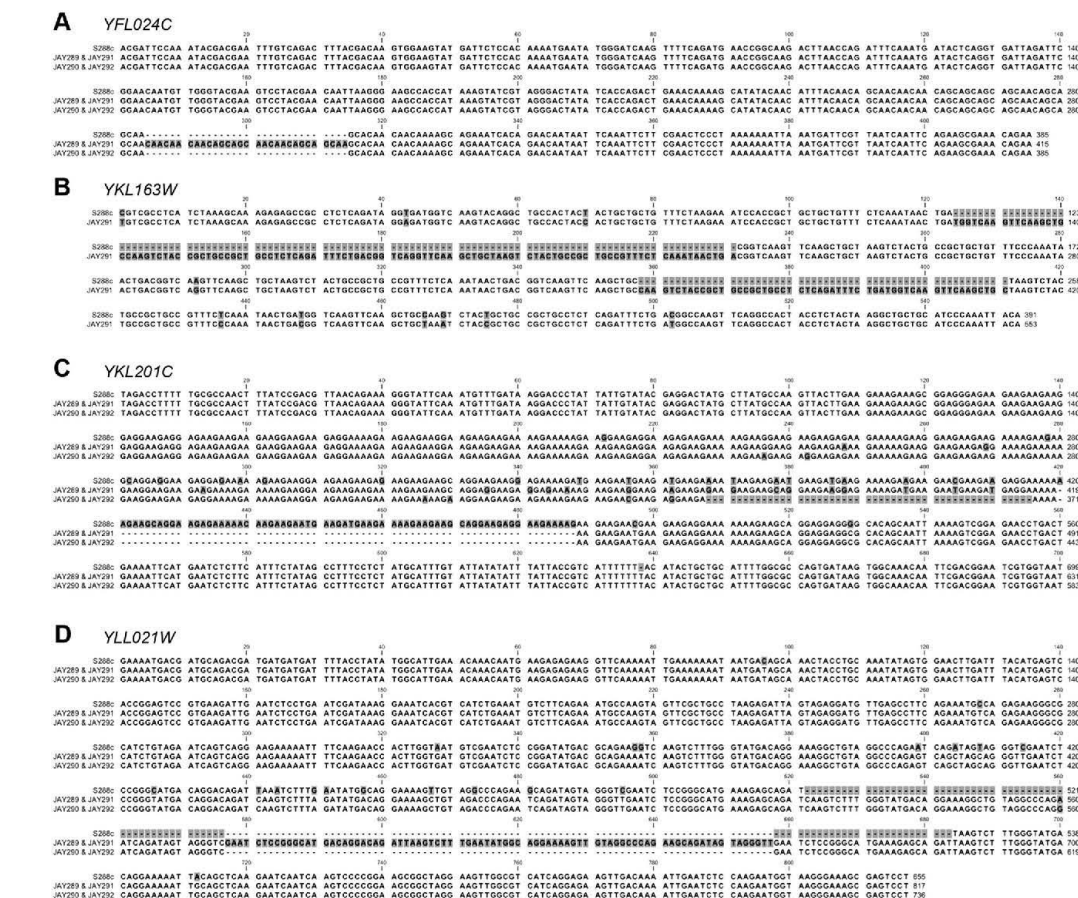


Fig. 2. Length polymorphisms within coding sequences.

Alignments of four of the genomic DNA sequences used as PCR marker loci in this study. Panels A through D show the alignments between the predicted PCR products (from Forward to Reverse primers) from the reference strain S288c and from JAY270/PE-2 sequences. Polymorphisms are shown shaded in gray, and insertion/deletion gaps are shown as dashes. For the loci where JAY270/PE-2 is homozygous (B) only the JAY291 haploid sequence is shown; and for the loci where this strain is heterozygous (A, C, D), the sequences of the two alleles are shown separately.

and JAY291 for all 44 genes described in Marinangeli et al. (2004) and in Verstrepen et al. (2005) (Marinangeli et al., 2004; Verstrepen et al., 2005). 16 of the 44 genes were either not fully assembled or did not have a strong similarity hit in JAY291. Therefore, only 28 of the 44 genes were discoverable using our approach. 17 of the 28 discoverable genes either had no size polymorphism, or had polymorphisms that were too small (at or below the 24 bp size cut off). Of the 11 remaining genes, 3 were successfully identified by our approach (above), but 8 were not. The reasons why some specific genes were missed by our approach are not entirely clear, but they include in a few cases a lack of size polymorphism relative to YJM789, RM11-1a, and AWRI1631 (Section 2). We compiled a list with all 50 genes from the three studies as a source of potentially useful PCR markers for future *S. cerevisiae* genetic diversity studies (Supplementary Table 2). We focused the rest of our study on the nine genes listed in Table 1.

In most cases, the polymorphic regions identified in our study contained low complexity sequences, often short repetitive minisatellites (Fig. 2; and Supplementary Figure 1). Others, such as YLL021W, had internal repeat units that were larger in size (81 bp).

In all cases, the length polymorphic regions were present as in-frame insertions or deletions in the S288c and JAY291 coding sequences, and none of the selected genes contained introns. Three of the genes (*YDR299W*, *YFL024C*, and *YPR143W*) are essential for viability in S288c according to the *S. cerevisiae* systematic deletion studies. The nine genes participate in assorted cellular pathways, including, but not limited to, membrane or cell wall proteins. The primary goal of this study was to use of these sequences as molecular strain identification markers, therefore we did not investigate any possible association between locus function and the presence of length polymorphisms.

Once the candidate length polymorphisms were identified, we designed oligonucleotide primer pairs flanking those sites to generate PCR products (Table 1). We used these primers to amplify genomic DNA from FY86 (an S288c isogenic strain; (Winston et al., 1995)) and from JAY291, and confirmed the expected product size polymorphisms in seven of the nine genes (Fig. 3A). For two of the genes, *YKL163W* and *YKL201C*, the resulting JAY291 PCR products had different sizes from those predicted from the published genomic sequence. However, in both cases the observed



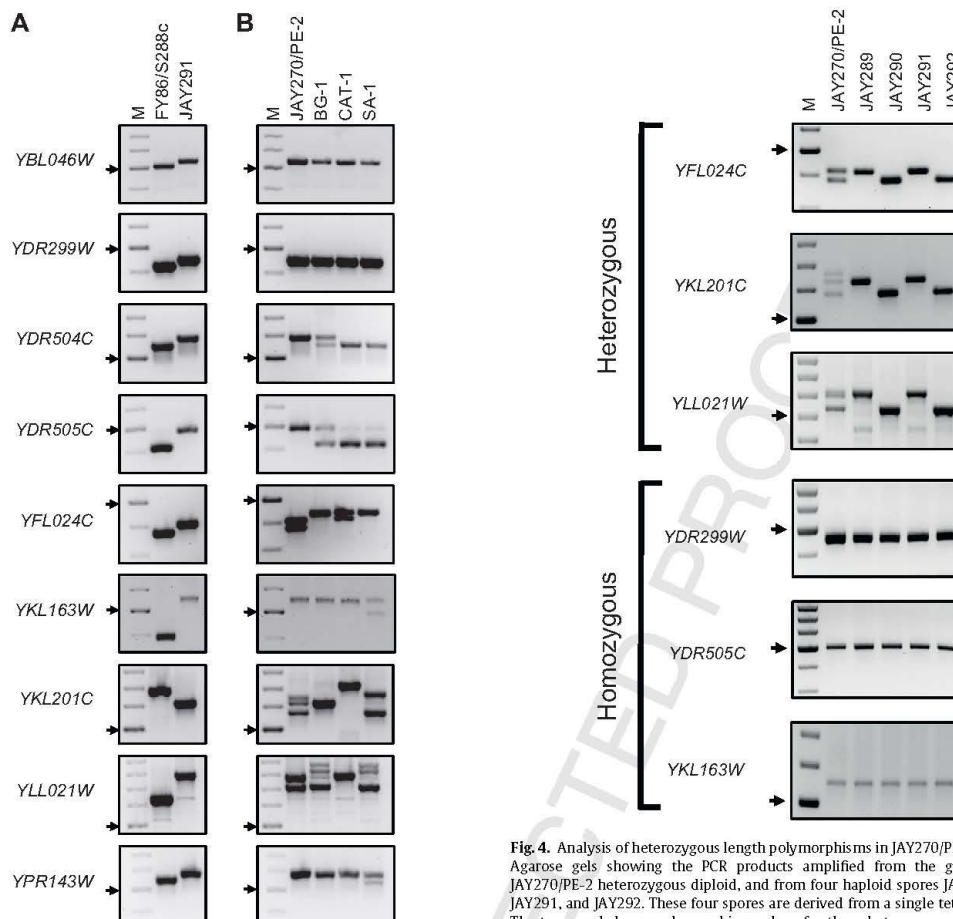


Fig. 3. PCR amplification of polymorphic loci.

Panel A shows agarose gels with PCR products amplified from the genomes of the reference strain FY86/S288c and of the JAY291 haploid. Panel B shows the PCR products obtained from the JAY270/PE-2 heterozygous diploid, and from three other widely used bioethanol yeast strains. M indicates the molecular weight ladder, and the arrow indicates the 500bp marker band. The only exception is for the YLL021W gel for which the 700bp band is indicated.

Fig. 4. Analysis of heterozygous length polymorphisms in JAY270/PE-2.

Agarose gels showing the PCR products amplified from the genome of the JAY270/PE-2 heterozygous diploid, and from four haploid spores JAY289, JAY290, JAY291, and JAY292. These four spores are derived from a single tetrad of JAY270. The top panel shows polymorphic markers for three heterozygous loci, with 2:2 Mendelian segregation among the haploids. The bottom panel shows the banding patterns for three homozygous loci. M indicates the molecular weight ladder, and the arrow indicates the 500bp marker band. The only exception is for the YLL021W gel for which the 700bp band is indicated.

bands were still polymorphic relative to S288c. The likely cause for this discrepancy was local error in the *de novo* assembly of the JAY291 genome (Argueso et al., 2009). Repetitive and low complexity regions are particularly challenging to assemble accurately from the short sequencing reads used to determine the JAY291 genomic sequence (mostly 35 bp and 50 bp Illumina GAII reads, with some 200 bp Roche 454). Therefore, we re-sequenced the JAY291 PCR products spanning the YKL163W and YKL201C polymorphic regions using long reads from conventional Sanger sequencing. The Sanger-determined sequences of both genes matched the sizes of the PCR products (Figs. 2 and 3A).

### 3.2. PCR identification of industrial bioethanol strains

Following the identification and validation of the intragenic length polymorphisms between S288c and JAY291, we used the same primer sets to look for polymorphisms between our strain of

interest JAY270/PE-2, and strains BG-1, CAT-1 and SA-1 (Fig. 3B). This group represents the four most widely adopted strains in the Brazilian bioethanol industry (Basso et al., 2008). This analysis revealed polymorphisms at seven of the nine loci assayed, confirming that our set of markers was highly informative for the identification bioethanol strains.

In several cases, more than one PCR product was amplified from genomic DNA of the industrial strains. We previously showed that JAY270/PE-2 is a heterozygous diploid, therefore suggesting that this strain, as well as the other industrial strains tested, may be heterozygous at the polymorphic loci. To test this hypothesis, we prepared genomic DNA from JAY270/PE-2 and from all four haploid spore products of a JAY270/PE-2 meiotic tetrad (Fig. 4). We used this material as template in PCR reactions to amplify three candidate homozygous loci (YDR299W, YDR505C, and YKL163W) and three candidate heterozygous loci (YFL024C, YKL201C, and YLL021W). The four haploid spores in this tetrad included JAY291 (the one with a sequenced genome), and its three sibling spores JAY289, JAY290, and JAY292. For the homozygous markers, we observed a single PCR band of the same size in all samples. In contrast, for the YFL024C and

**Table 2**

Analysis of PCR length polymorphic alleles in common sugarcane yeast strains.

Locus	YBL046W	YDR299W	YDR504C	YDR505C	YFL024C	YKL163W	YKL201C	YLL021W	YPR143W
PIC <sup>a</sup>	0.30	0.47	0.38	0.30	0.74	0.73	0.89	0.89	0.61
Strains <sup>b</sup>									
JAY270/PE-2	2	2	2	2	2/3	2	2/3/6	2/9	2
CAT-1	2	2	1	1	4	2	4	4	2
BG-1	2	2	1/2	1/2	4	2	6	3/8/10	2
SA-1	2	2	1	1	4	2/7	2/5	3/8/10	1/2
FY86/S288c	1	1	1	1	1	1	1	1	1
CA-11	2	2	3	1	4	7	2	1	1
IZ1904	2	2	1	2	3	2	5	5	1
Y904	1	3	1	1	1	8	1	3	2
M304-2C	1	1	1	1	1	8	1	1	1
FF	2	3	1	1	2	7/4	5/7	4/3	2
CL-A	2	1/2	1	1	2/3	4	6/7	4	3
SJC1 230/8A	2	2	1	1	2	2/7	2	2/9	1
SJC1	2	2	2	2	1	3	8/11	5	1
SJC3	2	2	1	1	1/2	2	11	2/9	1
SJC6	2	2	1	1	3	2/7	2/6/11	4	1/4
SJC7	2	1	1	1	1/3	2	3/9	6/7/10	1/4
PRGC4A	1/2	2	1	1	2	2/7	4/9	6/7	1
EQ287C1	2	2	1	1	2	5/6	10	7	1/4
EQ287C2	2	2	1	1	3/4	2	1	7	4
LQ1	2	2/3	1	1	1/3	2	1/4/9	3/7/10	1/2
LQ2	2	2	3	1	1/3	2/7/9	6/9/10	6/7	1

<sup>a</sup> Discriminatory power of each marker locus was estimated by the Polymorphic Index Content (PIC) parameter (Section 2).<sup>b</sup> The numbers identify the specific PCR allele detected in each strain. The allele observed in the FY86/S288c reference genome was arbitrarily assigned as allele "1". The allele observed in JAY291 was arbitrarily assigned as allele "2".

YLL021W candidate heterozygous loci we observed the amplification of two bands in the JAY270/PE-2 diploid, and a single band in each of the haploid spores. PCR at the YKL201C locus produced three bands in JAY270/PE-2: The two shorter bands corresponded to the alleles observed in the haploids; and the longer third band was a specific and reproducible chimeric product (characterization and discussion presented in Supplementary Figure 2). The polymorphic PCR products from all the three heterozygous loci displayed the expected 2:2 Mendelian segregation among the spores. Finally, we sequenced the second (unknown) alleles from YFL024C, YLL021W, and YKL201C, and compared them to the first alleles already known from the JAY291 genomic sequence (Fig. 2). In the case of YFL024C, the allele found in JAY290 and JAY292 was 385 bp long, and was the same as in S288c. For YLL021W, we observed that the allele present in JAY290 and JAY292 had two units of an 81 bp internal repeat, while the JAY289 and JAY291 allele has three units, and S288c has only one unit. Finally, in the case of YKL201C, the allele found in JAY290 and JAY292 was 583 bp long, and differed from both the 631 bp allele in JAY291 and JAY289, and also from the 699 bp allele

from S288c. This analysis showed that the heterozygosis observed in JAY270/PE-2 and in other industrial yeast strains is a helpful feature that further enhances the discriminatory power of our PCR markers.

To further assess the usefulness of the PCR markers we analyzed a larger panel of yeasts associated with sugarcane fermentation. These included all the strains described above, as well as a *Cachaça* still strain (CA-11), three minor bioethanol strains (IZ1904, Y904, M304-2C), eleven indigenous contaminant yeast strains isolated from Brazilian bioethanol and *Cachaça* fermentations, and a widely used baker's starter strain (FF). The PCR products were visualized in conventional agarose gel electrophoresis (Supplementary Figure 3), and the product sizes for each strain were examined to identify their specific allelic length variants. The summary of this analysis is shown in Table 2. We observed several new PCR alleles that were amplified in the different strain/loci combinations. In two cases, YBL046W and YDR505C, the only size polymorphism observed was the one originally mapped between S288c and JAY291. All other loci produced between three and eleven PCR products of discernible

**Table 3**

PCR analysis of yeast populations in six bioethanol distilleries during the 2011 season.

Strain	Distilleries																	
	Aldia			Morro Vermelho			Eldorado			C. do Pontal			Santa Luzia			Alto Taquari		
	Inoc. (tons) <sup>a</sup>	1 <sup>st</sup> b	2 <sup>nd</sup> c	Inoc. (tons)	1 <sup>st</sup>	2 <sup>nd</sup>	Inoc. (tons)	1 <sup>st</sup>	2 <sup>nd</sup>	Inoc. (tons)	1 <sup>st</sup>	2 <sup>nd</sup>	Inoc. (tons)	1 <sup>st</sup>	2 <sup>nd</sup>	Inoc. (tons)	1 <sup>st</sup>	2 <sup>nd</sup>
PE-2	0.35	22	18	0	0	0	0.1	12	19	0	0	0	0.3	0	0	10	10	0
CAT-1	0.35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	5	0
BG-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	2	0
Angest	0	0	0	4	0	0	1.5	0	0	2	1	0	2	0	0	0	0	0
FF	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Indigenous <sup>d</sup>	0	13	2	0	15	15	0	3	1	0	14	20	0	20	20	0	3	20
Total	0.7	35	20	4	15	15	1.6	15	20	2	15	20	4.3	20	20	30	20	20

This analysis only included the molecular markers for loci YFL024C, YKL163W, YKL201C and YLL021W. Colonies with a heterozygous, or a homozygous pattern for one of the two YFL024C, YKL201C or YLL021W alleles present in JAY270/PE-2 were classified as derived from the starter strains.

<sup>a</sup> Tons of dry active yeast powder used as starting inoculum (March/April).<sup>b</sup> Number of colonies identified at the first sample collection after the beginning of the season (June/July).<sup>c</sup> Number of colonies identified at the second sample collection (September/October).<sup>d</sup> Colonies whose genotype that did not match any of the other industrial strains in the starting inoculum were classified as indigenous.

Please cite this article in press as: Carvalho-Netto, O.V., et al., A simple and effective set of PCR-based molecular markers for the monitoring of the *Saccharomyces cerevisiae* cell population during bioethanol fermentation. J. Biotechnol. (2013), <http://dx.doi.org/10.1016/j.jbiotec.2013.08.025>



size. Four of the loci, *YFL024C*, *YKL163W*, *YKL201C*, and *YLL021W*, were particularly polymorphic, and we found that using only those loci provided sufficient discriminatory power to uniquely identify every one of the 21 strains in Table 2. For the remainder of our study, we focused our PCR genotyping analyses on those four core markers.

### 3.3. Monitoring of yeast populations in industrial fermentations

After completing the design and testing of our PCR markers, we evaluated their practical effectiveness by analyzing the yeast populations in samples collected from six Brazilian distilleries during the 2011 sugarcane harvest season. Each of these distilleries began their fermentations in March or April using known specific mixes of commercial dry active yeast powder as initial inoculum (Table 3). As is typical in Brazil, the starting inoculum was often composed of a large proportion of baker's yeast strains (e.g. FF, Angest), supplemented with smaller amounts of bioethanol-specific strains such as PE-2, CAT-1 and BG-1. During the first few weeks of cell recycling the baker's strains gradually disappear, while the bioethanol-specific strains and/or indigenous contaminants progressively take over the population. This dynamic has been demonstrated previously and it is a common feature of the Brazilian bioethanol production process (Basso et al., 2008; da Silva et al., 2005).

We collected cell samples from each distillery roughly at the end of the first and second trimesters of the harvest season: June/July and September/October, respectively. We prepared genomic DNA from about twenty yeast colonies isolated from each sample, and used them in PCR. The amplification patterns from these colonies were then compared to those of the strains known to have been originally present in the inoculum. It is important to note that the banding patterns in the industrial strain colonies re-isolated from the industry were identical to the patterns previously observed in the laboratory stocks, suggesting that the PCR markers were genetically stable through prolonged cell propagation and fermentation.

The banding patterns for the colonies isolated from the distilleries were analyzed to generate a snapshot of the yeast population at each time (Table 3). Even though the colony sample sizes were small, this initial field test indicated that the markers are quite effective. Using only the four core markers, we were able quickly classify the colonies as being either derived from one of the initial starter strains, or indigenous when their banding pattern did not match any of those present in the inoculum.

As expected, the starter baker's yeast strains were mostly out-competed by the time of the first sampling, and were not detected at all in the second. The bioethanol strains had varying degrees of success in establishing themselves as the dominant members of the yeast population. For example, PE-2 was very effective in colonizing the fermentation environment at the Alcídia and Eldorado distilleries, but was replaced by better-adapted indigenous contaminants at Alto Taquari. Our PCR data showed that the indigenous yeasts were represented by multiple strains, ranging from two and up to eleven unique PCR genotypes depending on the distillery. Interestingly, at the Alto Taquari distillery a specific indigenous strain was detected at a high frequency in the yeast population (~60%), yet the overall bioethanol yield in the distillery remained satisfactory (data not shown). This suggested that this strain is well adapted to the specific environmental conditions at this distillery, and therefore could be isolated for use as a starter strain in subsequent years.

### 3.4. PCR length polymorphisms in non-sugarcane strains

Our analysis of a panel of sugarcane-associated strains (Table 2) revealed a substantial degree of polymorphism for our PCR

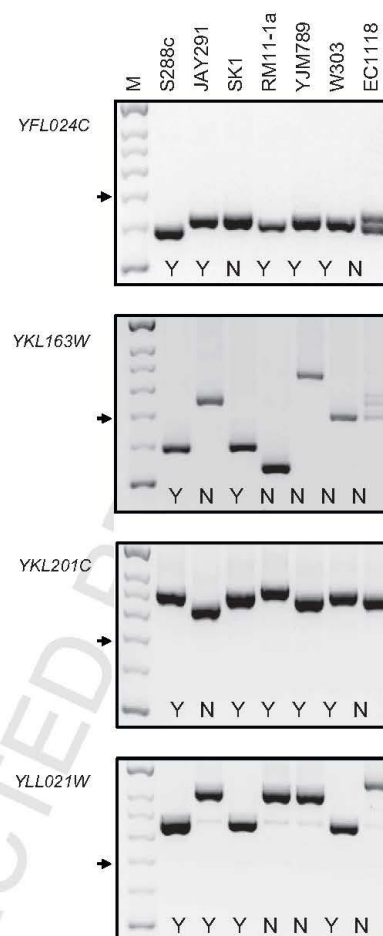


Fig. 5. PCR amplification of markers in *S. cerevisiae* strains unrelated to sugarcane fermentation.

The panels show the PCR products amplified from the indicated loci resolved in agarose gels. The yeast strains are listed above each lane. M indicates the molecular weight ladder, and the arrow indicates the 700 bp marker band. The letters within the panels indicate for each lane if the size of the observed products was either consistent (Y) with the size predicted from the deposited genome sequence of each strain, or not (N).

markers. In addition, our analysis of yeast samples collected at distilleries identified further allelic variants among indigenous contaminant strains (Table 3; and data not shown). Together these observations suggested that these loci are highly variable in structure, and that their usefulness as molecular markers may not be limited just to sugarcane strains. To assess the universal applicability of these markers, we took advantage of the several fully sequenced *S. cerevisiae* genomes available in public databases. We created BLASTn alignments using S288c as the reference sequence, and compared the sequence lengths at each locus in 59 strains from different origins (Borneman et al., 2008; Liti et al., 2009; Novo et al., 2009; Wei et al., 2007). This analysis, summarized in Supplementary Table 3, allowed us to uniquely differentiate all but two of the strains tested. These two strains (UWOP503.461.4 and

UWOPS05.227.2) were isolated from the same geographic region, and are very similar at the genomic level (Liti et al., 2009).

For most loci, the sequence alignments revealed a higher number of length variant alleles than we were able to observe in the sugarcane strain set (Table 2). This number ranged from only two alleles in YBL046W, up to seventeen alleles in YKL201C. We found cases in which the size difference between the predicted PCR products for specific alleles was quite large and easily detectable in conventional agarose electrophoresis gels (e.g. 240 bp, 192 bp, 164 bp), but also observed very small size differences (<15 bp) that would require running the PCR products on higher resolution gels.

While the alignments suggested that loci such as YKL201C are highly polymorphic (17 alleles), the interpretation of the results should be taken with caution since the determination of the genomic sequences at these low complexity regions is susceptible to errors associated with the mis-assembly of sequencing reads. We therefore compared the sizes of the PCR products predicted from the deposited genome sequences of S288c, JAY291, and five additional strains, to the actual products observed after amplification at the four core loci (Fig. 5). The correlation between the predicted and observed PCR product sizes was quite poor. S288c, which arguably has the most accurate *S. cerevisiae* genome assembly, was the only strain that matched the size prediction for all four loci tested. The other six strains combined had only 12 size matches in 24 alleles tested. This result underscored the difficulty of accurately assembling low complexity regions, even in the case of genomes such as YJM789 and RM11-1a that were sequenced primarily with conventional Sanger reads.

Despite the limitations caused by genome mis-assembly, taken together, the *in silico* PCR analysis and the validation experiment confirmed that the markers are indeed polymorphic in strains outside the sugarcane set, even if the predicted size did not frequently correspond to the observed PCR product. The results also showed that in a reasonable number of cases the size difference between polymorphic PCR products was sufficiently large to be resolved in conventional agarose electrophoresis. Therefore the PCR markers developed in this study should be universally applicable to other fermentation processes that require a simple and fast method for monitoring the yeast population.

#### 4. Conclusions

Previously described PCR-based methods for the identification of yeast strains relied primarily on DNA sequences predicted to be variable in the genome of the laboratory strain S288c. In contrast, the strategy we adopted in this study was to directly identify relatively large length polymorphisms in our industrial strain of interest, and then design PCR markers at those regions. This approach allowed us to create a simple and effective system for positively identifying yeast colonies derived from JAY270/PE-2 in the heterogeneous yeast populations found in bioethanol distilleries. We showed that this marker set was also informative for other strains associated with sugarcane production, and possibly in strains from other origins. The PCR markers described here represent an attractive alternative for bioethanol distilleries, that may elect to carry out this simple and fast procedure in their on-site quality control laboratories, rather than relying on a third party for the analysis the yeast population.

#### Competing interests

The authors report no conflicts of interest.

#### Author's contributions

All authors contributed to the data analysis, and read and approved the final manuscript. O.V.C.N designed the concept, performed experiments and wrote the manuscript; M.F.C. and G.G.L.C performed bioinformatics analysis to identify polymorphic regions; A.R. and W.O.B. performed PCR genotyping analysis; J.L.A. designed the concept, performed experiments and wrote the manuscript; G.A.G.P. wrote the manuscript.

#### Endnotes

The use of the marker loci described in this work for commercial/for profit yeast genotyping purposes is subject to intellectual property protection through a pending patent application submitted to the Brazilian National Institute of Intellectual Property (Instituto Nacional de Propriedade Intelectual - INPI); patent application number 018100029693, 2010.

#### Acknowledgements

We would like to thank Odebrecht Agroindustrial and Brazilian funding agencies FAPESP and CNPq for partial funding support to this project and Graça Stupiello, who generously provided yeast strains. Bioenergy research in the Argueso laboratory was supported in part by a seed grant from the Colorado State University Clean Energy Supercluster.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jbiotec.2013.08.025>.

#### References

- Amorim, H.V., Lopes, M.L., de Castro Oliveira, J.V., Buckneridge, M.S., Goldman, G.H., 2011. Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 91, 1267–1275.
- Argueso, J.L., Carazzolle, M.F., Mieczkowski, P.A., Duarte, F.M., Netto, O.V.C., Missawa, S.K., Galzerani, F., Costa, G.G.L., Vidal, R.O., Noronha, M.F., Dominska, M., Andrietta, M.G.S., Andrietta, S.R., Cunha, A.F., Gomes, L.H., Tavares, F.C.A., Alcarde, A.R., Dietrich, F.S., McCusker, J.H., Petes, T.D., Pereira, G.A.G., 2009. Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19, 2258–2270.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, A.J., Struhl, K., 1998. *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
- Basso, L.C., de Amorim, H.V., de Oliveira, A.J., Lopes, M.L., 2008. Yeast selection for fuel ethanol production in Brazil. *Fems Yeast Res.* 8, 1155–1163.
- Borneman, A.R., Forgan, A.H., Pretorius, I.S., Chambers, P.J., 2008. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *Fems Yeast Res.* 8, 1185–1195.
- Bovo, B., Andrighetto, C., Carlot, M., Corich, V., Lombardi, A., Giacomini, A., 2009. Yeast population dynamics during pilot-scale storage of grape marcs for the production of Grappa, a traditional Italian alcoholic beverage. *Int. J. Food Microbiol.* 129, 221–228.
- Campos, C.R., Silva, C.F., Dias, D.R., Basso, L.C., Amorim, H.V., Schwan, R.F., 2010. Features of *Saccharomyces cerevisiae* as a culture starter for the production of the distilled sugar cane beverage cachaca in Brazil. *J. Appl. Microbiol.* 108, 1871–1879.
- da Silva, E.A., dos Santos, S.K.B., Resende, A.D.M., de Moraes, J.O.F., de Moraes, M.A., Simoes, D.A., 2005. Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting. *Anton Leeuw Int J G* 88, 13–23.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 genes. *Science* 274, 546–547–563.
- Gomes, L.H., Duarte, K.M.R., Argueso, J.L., Echeverrigaray, S., Tavares, F.C.A., 2000. Methods for yeast characterization from industrial products. *Food Microbiol.* 17, 217–223.
- Hierro, N., Esteve-Zarzoso, B., Gonzalez, A., Mas, A., Guillamon, J.M., 2006. Real-time quantitative PCR (QPCR) and reverse transcription-QPCR for detection and enumeration of total yeasts in wine. *Appl. Environ. Microb.* 72, 7148–7155.

Please cite this article in press as: Carvalho-Netto, O.V., et al., A simple and effective set of PCR-based molecular markers for the monitoring of the *Saccharomyces cerevisiae* cell population during bioethanol fermentation. *J. Biotechnol.* (2013), <http://dx.doi.org/10.1016/j.jbiotec.2013.08.025>



- Kane, S.M., Roth, R., 1974. Carbohydrate metabolism during ascospore development in yeast. *J. Bacteriol.* 118, 8–14.
- Kurtzman, C.P., Robnett, C.J., 1998. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Anton Leeuw Int. J. G* 73, 331–371.
- Legras, J.L., Karst, F., 2003. Optimization of interdelta analysis for *Saccharomyces cerevisiae* strain characterization. *FEMS Microbiol. Lett.* 221, 249–255.
- Legras, J.L., Ruh, O., Merdinoglu, D., Karst, F., 2005. Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *Int. J. Food Microbiol.* 102, 73–83.
- Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O'Kelly, M.J.T., van Oudenaarden, A., Barton, D.B.H., Bailes, E., Ba, A.N.N., Jones, M., Quail, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., Louis, E.J., 2009. Population genomics of domestic and wild yeasts. *Nature* 458, 337–341.
- Marinangeli, P., Angelozzi, D., Ciani, M., Clementi, F., Mannazzu, I., 2004. Minisatellites in *Saccharomyces cerevisiae* genes encoding cell wall proteins: a new way towards wine strain characterisation. *Fems Yeast Res.* 4, 427–435.
- Martorell, P., Querol, A., Fernandez-Espinar, M.T., 2005. Rapid identification and enumeration of *Saccharomyces cerevisiae* cells in wine by real-time PCR. *Appl. Environ. Microb.* 71, 6823–6830.
- Mortimer, R.K., Romano, P., Suzzi, G., Polsinelli, M., 1994. Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts. *Yeast* 10, 1543–1552.
- Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.L., Wincker, P., Casaregola, S., Dequin, S., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16333–16338.
- Querol, A., Barrio, E., Huerta, T., Ramon, D., 1992. Molecular monitoring of wine fermentations conducted by active dry yeast strains. *Appl. Environ. Microb.* 58, 2948–2953.
- Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.
- Schuller, D., Pereira, L., Alves, H., Cambon, B., Dequin, S., Casal, M., 2007. Genetic characterization of commercial *Saccharomyces cerevisiae* isolates recovered from vineyard environments. *Yeast* 24, 625–636.
- Schuller, D., Valero, E., Dequin, S., Casal, M., 2004. Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiol. Lett.* 231, 19–26.
- Slater, G.S., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6.
- Smith, J.S.C., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S., Ziegler, J., 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zeamays* L.): comparisons with data from RFLPS and pedigree. *Theor. Appl. Genet.* 95, 163–173.
- Vaudano, E., Garcia-Moruno, E., 2008. Discrimination of *Saccharomyces cerevisiae* wine strains using microsatellite multiplex PCR and band pattern analysis. *Food Microbiol.* 25, 56–64.
- Verstrepen, K.J., Jansen, A., Lewitter, F., Fink, G.R., 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990.
- Wei, W., McCusker, J.H., Hyman, R.W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M., Wilhelmy, J., Komp, C., Tamse, R., Wang, X., Jia, P., Luedi, P., Oefner, P.J., David, L., Dietrich, F.S., Li, Y., Davis, R.W., Steinmetz, L.M., 2007. Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12825–12830.
- Wheals, A.E., Basso, L.C., Alves, D.M.G., Amorim, H.V., 1999. Fuel ethanol after 25 years. *Trends Biotechnol.* 17, 482–487.
- Winston, F., Dollard, C., Ricupero, S.L., 1995. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288c. *Yeast* 11, 53–55.



## **Anexo II**



## DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação/tese de Doutorado intitulada "Identificação e caracterização de genes e fatores relacionados à floculação e desenvolvimento de uma levedura industrial não floculante":

( ) não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

Tem autorização da(s) seguinte(s) Comissão(ões):

( X ) CIBio – Comissão Interna de Biossegurança , projeto No. projeto No. 2011/03, Instituição: Universidade Estadual de Campinas

( ) CEUA – Comissão de Ética no Uso de Animais , projeto No. \_\_\_\_\_, Instituição: \_\_\_\_\_

( ) CEP - Comissão de Ética em Pesquisa, protocolo No. \_\_\_\_\_, Instituição: \_\_\_\_\_

*\* Caso a Comissão seja externa ao IB/UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*

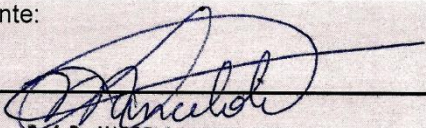
  
Aluna: Aline Rodrigues

  
Orientador: Gonçalo Amarante Guimarães Pereira

Para uso da Comissão ou Comitê pertinente:

(X) Deferido ( ) Indeferido

Carimbo e assinatura

  
Prof. Dr. MARCELO LANCELOTTI  
Presidente da Comissão Interna de Biossegurança  
Instituto de Biologia - UNICAMP

Para uso da Comissão ou Comitê pertinente:

( ) Deferido ( ) Indeferido

Carimbo e assinatura